

SPECIFIC AIMS

Our overall goal is to establish a new empirical paradigm to unravel the functional architecture of complex genetic disorders, with broadly reaching implications. The specific problem to be addressed is the failure of genome wide association studies (GWAS) to fully explain the heritability of complex genetic disorders. Given the immense complexity of the biological networks that underlie health and disease, this frailty is hardly surprising and it underscores the need for new empirically testable approaches to relate GWA genes to disease causality. Our approach to tackle this fundamental issue is guided by the following tenets: i) that complex genetic disorders arise in poorly buffered biological networks, ones that are especially sensitive to changes in gene dosage or protein function; ii) that to first approximation genes cataloged in GWAS are predominantly partial loss or gain of function alleles; iii) that these alleles can be modeled by haploinsufficiency studies – comparing the effects of orthologous genes in the hemizygous states in model systems; and iv) that the analysis of haploinsufficiency in a combinatorial manner, termed *complex haploinsufficiency* (CHI), will map the genetic interaction networks that are most fragile and therefore critical for each disease.

With these concepts in mind, we propose to exploit CHI as a novel and potentially revolutionary approach to model the functional architecture of complex genetic disorders. To deliver in a timely manner, we will take advantage of the wealth of currently available and emerging gene knock-out (KO) resources. These resources will be configured for CHI analysis of disease pathways in a manner that promises to reveal highly ordered and interpretable genetic infrastructures.

Our overall hypothesis is that analyzing CHI will identify the molecular pathways and interactions that, when incomplete, account for the missing heritability in GWAS.

We propose the following Specific Aims:

1: Establish the principles of CHI in diploid yeast. Diploid yeast will be used to operationally define CHI in this uniquely qualified unicellular model system. We will select approximately 200 genes from four functional classes and measure fitness for all single and double hemizygous strains under a variety of chemical perturbations. Our hypothesis is that genes that function in vital biological processes will exhibit dense patterns of CHI because the organism is especially sensitive to defects in these pathways.

2: Apply CHI to define the functional genetic architecture in mouse models of systemic lupus erythematosus (SLE) and coronary artery disease (CAD). We will take advantage of a toll-like receptor 7 (*Tlr7*) duplication to genetically sensitize mice for SLE (**Aim 2a**), and an atherogenic diet to environmentally sensitize mice for CAD (**Aim 2b**). We will configure the genetics based on orthologs of GWAS-associated genes that permit the evaluation of simple and complex haploinsufficiencies. In both cases, a robust combination of disease phenotype biomarkers will be quantified. Finally, whole genome gene expression (GE) analysis of relevant tissues will provide the information needed for molecular pathway inferences.

3: Develop computational algorithms that model the functionally relevant mammalian genetic architecture revealed by CHI. We will derive interactive genetic networks that map the disease phenotypes to the underlying molecular pathways of SLE and CAD based on mouse orthologs of GWA genes. Our hypothesis is that orthologs and combinations thereof that are most sensitive to dosage variation will identify molecular pathways and interactions that better explain the missing heritability in GWAS.

4: Determine whether CHI patterns revealed are predictive of disease and its severity in a genetically diverse population. The preceding Aims model disease-relevant molecular networks when the genetic flux is limited to dosage variation in single and paired GWAS ortholog combinations. Our goal here is to evaluate the durability of the identified disease networks in a more dynamic genetic context. We will use the SLE and CAD models derived in Aim 3 to predict the phenotypic outcomes for more than two genetic perturbations. To evaluate these predictions, we will employ a mouse outcross strategy to vary gene dosage of GWAS orthologs in a highly permutational manner to address the genetic fragility and robustness of disease networks in a context that is more akin to genetic variation at the population level. Our hypothesis is that successful predictions will identify correctly inferred networks and failed predictions will identify incomplete networks for a second iteration of computational modeling.

RESEARCH STRATEGY

CHALLENGE, INNOVATION, AND IMPACT STATEMENT

The recent proliferation of genome-wide association studies (GWAS) has provided an unprecedented view of the genetics of the most common complex human diseases [2]. However, the genetic variants identified have proven difficult to validate and each tend to account for only a small percentage of the estimated heritable risk [3, 4]. New methods to formulate and test hypothesis from GWAS are therefore needed. Our approach uses unique assets of the laboratory mouse to establish a new paradigm for disease models. It configures currently available mouse knockout alleles to model partial loss or gain of functional alleles of GWA orthologs. By considering the combined effects of these alleles, termed complex haploinsufficiency, we will derive genetic interaction networks that reveal fundamental principles of how subtle genetic perturbations of fragile pathways lead to disease in model systems. The proposed work will serve as a template for studying how multiple GWA genes form functional networks that influence disease. It will enable quantitative hypothesis generation and empirical testing for the wealth of genetic data now available. With current and future resources for mouse genetics such as the knockout mouse project (KOMP), this approach will be widely applicable to human diseases that can be modeled in the mouse.

RATIONALE

The principles that underlie the proposed studies are evolved behaviors of eukaryotic systems and their underlying biological networks. Our reasoning is that these principles can be used to reveal the functional architecture of complex genetic diseases in both humans and mouse models. More specifically, we propose that a major source of the missing heritability in GWAS is the robustness inherent to biological systems. Genetic buffering, a general term for compensatory mechanisms maintain homeostasis in the event of one or more genetic perturbations, limits the efficacy of the single-locus association strategies commonly used. At a more fundamental level, these buffering mechanism likely limit disease development to pathways that are unusually fragile with respect to genetic perturbations. While these biological constraints present major hurdles to GWAS-centered approaches, we argue that they can be addressed through our novel application of the mouse model system. Genes cataloged in GWAS can be generally considered as either partial loss or gain of function alleles – variants that, regardless of their nature, weaken or strengthen their respective proteins' functions. The essence of these variants can be modeled conveniently in mice by taking advantage of available knock-out (KO) alleles to reduce gene dosage to the hemizygous state. Mouse orthologs of GWA genes are significantly enriched in haploinsufficient phenotypes provides evidence that these orthologs lie in relatively fragile pathways (discussed below). Therefore dosage perturbations of GWAS orthologs, readily produced using KO alleles, can be used to model the fragile pathways that lead to disease. Finally, by combining these variants in combinations of complex haploinsufficiency (CHI), it will be possible to interconnect genetic interactions that are most dosage sensitive. These perturbations can be readily combined to map networks of CHI in an efficient manner. This overall design provides a new approach to understanding complex diseases for which multiple weak-effect alleles have been identified via GWAS.

APPROACH

Conceptual Drivers of the Approach

Complex genetic disorders arise in especially fragile biological networks. We propose that the observed weakness in GWAS is due to the robust genetic architecture of biological systems. In response to fitness challenges and genetic mutations, organisms have evolved functional networks that buffer phenotypes against internal and external perturbations [5-7]. With this view of biology, organisms are evolved to maintain homeostasis and diseases are the result of relatively rare failures of biological networks that are insufficiently buffered against perturbation. Disease states are due to disregulation that occurs in exceptionally fragile pathways or networks. Furthermore, these states often require deleterious mutations in two or more genes to significantly affect pathway function [8]. Therefore the multiple genetic variants underlying disregulation will necessarily be obscured and often partially buffered by the complex network within which they are embedded. Taking this view, we propose that the small effect sizes attributable to GWA genes are a consequence of those genes operating in a network of genetic variance, which has a combined effect far greater than can be statistically attributed to any individual gene.

The idea that disease results from pathway fragility presents a challenge in understanding the underlying biology of the disease and its treatment. Identifying the genes that interact and mapping their network function will prove difficult if the network has a degree of built-in resistance to the perturbation of those genes. Predicting the effects of candidate therapeutic agents becomes difficult when those effects are filtered and processed through a robust biological network. Hypotheses generated inferentially by GWAS are difficult to validate due to the experimental inaccessibility of humans. We propose, instead, a research strategy using the empirical assets of laboratory mice specifically focused on the complicated biology underlying GWA genes, taking into account the robust features of biological systems and the essential nature of the functional variation in disease-related alleles.

The essence of the functional allelic variation underlying GWAS can be modeled by controlling gene dosages. The central theme of our approach is to combine dosage variants of GWAS orthologs in mice to study the biological networks that underlie the observed associations. The motivation is that the effects of GWA candidate genes have proven to be difficult to interpret, even when a candidate SNP is identified. While “rare” variants will potentially account for more of the genetic risk [9], the vast majority of associated SNPs (over 80%) do not lie in protein-coding regions of the genome [2]. Moreover, recent studies have shown that trait-associated SNPs are more likely to produce variations in transcript levels rather than in protein function [10-13]. Copy-number variants (CNVs) provide additional evidence to support the importance of gene dosage [14-17]. However, regardless of the nature of the polymorphism, they can be viewed as hypomorphic alleles that weaken or hypermorphic alleles that ultimately strengthen their proteins’ functions. We propose that the essence of the functionally relevant polymorphisms that underlie GWAS can be modeled by controlling gene dosages.

Approaches to model human diseases in mice. The laboratory mouse will continue to be the premier mammalian organism to model the impact of genetic variation on *in vivo* mammalian biology. A major asset is the ability to maximize the genetic contributions while minimizing the influence of environmental variables. Genetic dissections of mouse strains with natural genetic predisposition for diseases that mimic those developed in humans have lead to invaluable insights. However, they are best suited for gene discovery based on the genetic variation of mice, and only tangentially relate to human diseases. Gene-specific approaches using conventional gene knockouts have been similarly invaluable, but they are commonly used in the homozygous null state and do not model the more subtle forms of allelic variation that confer risk in humans. Recapitulating human allelic variants in mice by genetic knock-in or transgenic technologies is a promising bridge to model human diseases in mice. However, this approach is very labor and resource intensive, fraught with uncertainties, and can be practically applied to only a fraction of candidate disease-associated genes. Our approach leverages existing and forthcoming mouse knock-out resources and provides an efficient strategy for modeling the variable dosage effects of naturally occurring alleles using hemizygous alleles.

Mouse orthologs of GWA genes are unusually dosage sensitive. If GWA genes lie in pathways that are unusually sensitive to gene dosage variation, we would expect that they would express phenotypes when

reduced to the hemizygous state [18]. To test this hypothesis, we assembled a list of mouse knockouts and a sublist with known haploinsufficient phenotypes from the Mouse Genome Informatics (MGI) database (informatics.jax.org). Of 5222 unique knockout genes, 1350 expressed haploinsufficiency phenotypes. If GWA genes and their mouse orthologs lie in fragile pathways, we would expect for them to be enriched in haploinsufficient phenotypes. Of 1024 GWA genes with a mouse ortholog, 326 showed a hemizygous phenotype, significantly more than expected by chance ($p = 1.0 \times 10^{-6}$, Fisher's exact test). Although we recognize that the data contain errors of incompleteness at all stages, this enrichment supports our hypothesis that GWA genes lie in fragile pathways and networks that are most sensitive to gene dosage variation.

Complex Haploinsufficiency is a most promising strategy for modeling the functional architecture of complex genetic disorders. We propose that combining hemizygotes to study complex haploinsufficiency (CHI) will increase power to reveal weakly-buffered pathways. Although CHI has not been screened on a large scale in any organism, there are many examples of CHI in model organisms. A screen for CHI in yeast identified multiple genes that interact with actin [19]. A similar strategy of combining hemizygotes with small molecule screens identified many cases of synergistic effects [20]. CHI has been used to understand fly models of signaling and development, including the identification of genes involved in body segmentation [21], tyrosine kinase regulation [22], notch signaling and eye development [23, 24], and embryonic pair-rule segmentation [25].

The Notch/Jagged signaling pathway is a particularly instructive example. The exquisite sensitivity of this pathway to haploinsufficiency is documented in *Drosophila* [26], mice [27, 28], and in humans with Alagille syndrome [29, 30]. The principle of non-allelic non-complementation in the Notch pathway is exemplified in mice by the fact that *Jag1/Notch2* double hemizygous mice develop a disease that closely parallels Alagille syndrome while haploinsufficiency in *Notch2* and *Jag1* alone do not [31, 32].

With these principles in mind, as documented in multiple eukaryotic model organisms and humans, we note three important points that encapsulate the concept of our approach. First, finely tuned regulatory pathways are most susceptible to dosage effects caused by forms of allelic variation that weaken or strengthen the pathway. Second, dosage change of one component of the pathway sensitizes the pathway to dosage effects of other genes by mechanisms akin to CHI. Third, dosage sensitivity has proven to be a powerful screen to identify interacting components of regulatory pathways. Taken together, we propose that systematic studies of CHI in mouse models of disease will yield more complete models of complex genetic etiology.

Specific Aim 1: Establish the principles of CHI in diploid yeast.

The fundamental concept of this proposal is that disease susceptibility is often the result of multiple genetic defects with weak individual effects but a major cumulative effect. The experimental power unique to yeast provides the most direct means to address this concept by the study CHI on a large scale, and it will serve as a template for the mammalian studies to follow. Our hypothesis is that genes that function in vital biological processes will exhibit dense patterns of CHI because the organism is especially sensitive to defects in these pathways.

Synthetic genome array (SGA) technology [33] is a proven high-throughput technology to quantify and analyze genetic interactions in haploid yeast strains. To date, this method has been applied to haploid yeast. Fitness of over five million gene pairs have been measured, constituting a third of pair-wise gene deletions [8]. Double-deletion growth rates are quantified and compared to the growth rates of the two single deletions in isolation. Genetic interactions are identified as pairs for which the measured effect on growth rate is significantly different than the expected product of the two single-mutant effects. The 170,000 interactions detected provide a global map of genetic interactions. Pair-wise genetic buffering is manifest in this interaction network as negative genetic interactions, in which the growth defect exhibited by the double mutant is significantly more than the product of the two corresponding single-deletion growth defects. The importance of buffering is underscored by the fact that twice as many negative interactions were detected as positive interactions. Moreover, negative interactions were more enriched in common Gene Ontology (GO) annotations for interacting gene pairs. For example, genes involved in cell cycle regulation tended to exhibit negative interactions rather than positive. This is a process with multiple checkpoints that exhibits greater

Media	Targeted Gene Class
Rich media	None
MMS	DNA repair
Oleic Acid	Fatty acid metabolism
Salt	Stress and signaling
Erodoxin	Protein folding

Table 1 Media and gene classes to be tested in high-throughput CHI screen.

fragility when multiple genes are perturbed. In contrast, positive interactions are instances of genes combining to produce a less deleterious effect than expected. These can correspond to compensatory interactions in which one mutation masks another, and are similarly expected at a lower frequency in our CHI studies.

We propose to exploit this powerful system to assess CHI in diploid yeast on a genomic scale by carrying out similar experiments with double-hemizygote diploid strains. This study will be in collaboration with Dr. Brenda Andrews at the University of Toronto (see Letter of Collaboration). The diploid strains directly created by mating single-knockout haploid pairs, which is a common intermediate step in creating double-deletion haploids. We will select approximately 200 genes, evenly divided between four functional categories as defined by GO annotations. We will assay all strains for growth rate, a standard proxy for fitness in yeast assays, on different types of media. The categories are DNA repair, stress/signaling, peroxisome, and protein folding genes. We will first assess the frequency of CHI in growth on YPD. We will then assess CHI on media that targets each of the four gene classes (Table 1).

Results and expected outcomes

Our hypothesis is that CHI will be enhanced between genes with critical functions that respond to specific environments. We propose networks of function-specific CHI genes underlie GWAS findings, but marginal statistical power and population structure limit detection to a handful of genes and very few interactions. Yeast SGA screens will allow us to test this general concept on a large scale with well-characterized genes and controlled environments. By screening the genes for CHI on different media, we expect to find CHI is increased between the subset of genes that become important in that environment (Figure 1). This would

demonstrate that CHI occurs and that it is enriched in functions that are most important for survival. In the context of GWAS, this corresponds to disease susceptibility in the presence of specific environmental factors. For example, CHI in fatty acid metabolism genes is

expected to occur more frequently on media with oleic acid as a carbon source. This is conceptually similar to our proposed mouse study (Aim 2) in which dietary factors are expected to sensitize mice to heart disease caused by multiple genetic dosage defects in fragile pathways. In general, these studies in yeast will test our overall hypothesis and allow us to calibrate expectations in mouse model systems by providing a large-scale view of CHI.

Potential pitfalls and solutions

We foresee no technical limitations in creating and assaying the yeast strains. Double-hemizygote strains are a common intermediate step in SGA experiments and therefore have been previously performed on a large scale. The Toronto group has been performing SGA experiments and analysis for a number of years. Their statistical analysis techniques are able to quantify effects to high resolution, and we will consult them for questions arising about imaging analysis and growth quantification.

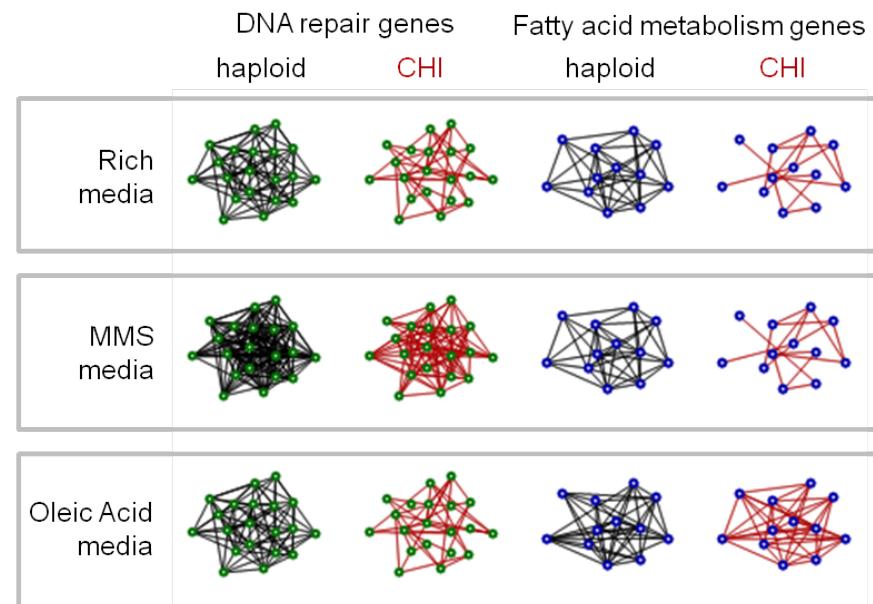
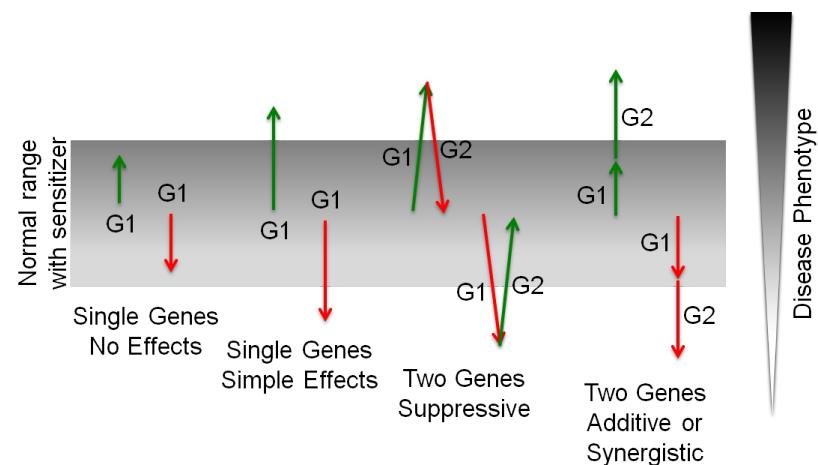


Figure 1. Expected haploid and complex haploinsufficiency (CHI) genetic interaction networks for different media. Subnetworks of DNA repair genes (green) are expected to exhibit greater interaction frequencies on DNA damaging media, especially for CHI (red). Similarly, fatty acid metabolism genes (blue) should exhibit more CHI when colonies are grown on oleic acid.

Specific Aim 2: Apply CHI to define the functional genetic architecture of mouse models for SLE (Aim 2a) and CAD (Aim 2b).

This Aim will evaluate the phenotypic effects of gene dosage variation in two established mouse models for human disease. The two models feature pathogenic processes that mimic SLE and CAD, respectively, and the disease genetics are highly orthologous to GWAS findings. Both disease processes can be readily measured by a robust series of quantitative biomarkers. Moreover, both diseases develop gradually, making it possible to measure the effects of gene dosage on accelerating or suppressing the disease as illustrated for simple and complex haploinsufficiency in Figure 2.

Figure 2. Schematic examples of gene dosage effects on disease. Normal range indicates the phenotypic range for wild-type mice with sensitizer (Yaa mutation for SLE, high fat diet for CAD). Dosage variation in two genes (G1 and G2) can have positive (green) or negative (red) effects. Single-gene effects can lie within the normal range or outside the normal range. Combinatorial effects of suppression can rescue phenotype while additive or synergistic effects compound phenotype.



Aim 2a. Determine the impact of gene dosage changes of orthologs of human SLE-associated genes on mouse SLE-like disease.

SLE is a common autoimmune disorder that involves the major innate and adaptive cells of the immune system. SLE characterized by autoantibodies and immune complexes that inflict systemic organ damage, often including the kidneys. It is a prototypic complex genetic disorder. Currently ~35 loci, often shared with other autoimmune disorders, contribute to its genetic etiology, and presumed environmental stimuli may be triggers [1, 3, 34-39]. This disease is a classic example of missing heritability in that the risk loci identified to date only explain ~15% of its genetic risk [3].

Consistent with the concept of fragile disease networks, multiple lines of evidence support the importance of dosage sensitivity in SLE [36, 40-45].

We will apply CHI of GWA orthologs to unravel the functional genomic architecture of SLE. At the molecular level, both GWAS and functional studies reveal and interconnect the Toll-like receptor (TLR) and the Type I interferon (IFN1) response pathways as a major functional module of SLE. Indeed, over 50% of the genes with risk for SLE map to the TLR/IFN1 response pathways [46]. Our gene selections will therefore be heavily based on these pathways. To insure that these key pathways are evoked, an integral part of our design is to genetically sensitize C57BL6/J (B6) mice for SLE by doubling the gene dosage of *Tlr7*. This is readily accomplished by incorporating of carrying the Y chromosome-linked autoimmune accelerator mutation, Yaa [47], which carries expressed copies of *Tlr7* on both sex chromosomes [42-44]. B6 mice carrying this develop an easily monitored, chronic form of SLE, with multiple IFN1 response and cytokine genes

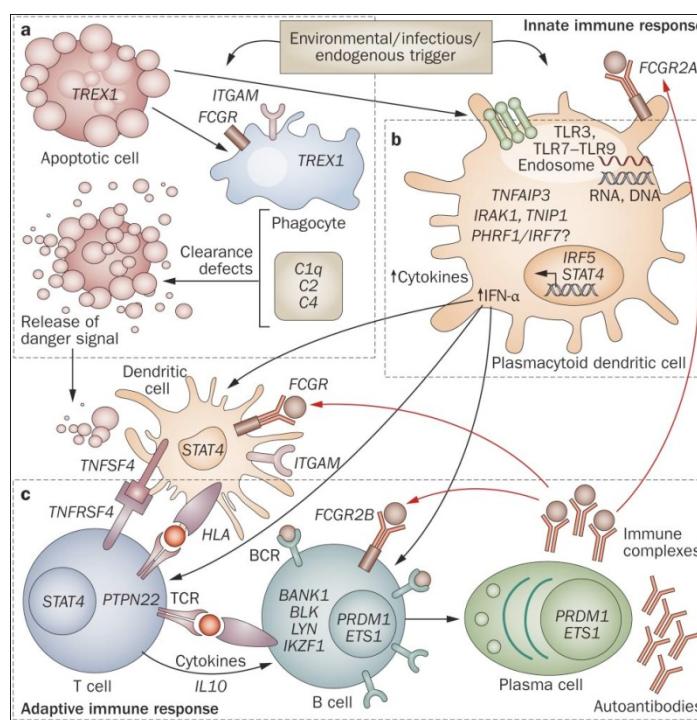


Figure 3. Model of SLE pathophysiology in the context of its genes with associated genetic risk and illustrating its complexity (from Ref. [1]).

upregulated as early as 6 weeks of age (new unpublished results). Moreover, this chronic disease can be amplified or suppressed by autosomal loci [42-44]. Thus, the expression of autoimmune disease in *Yaa* mice as measured by the biomarkers to be employed is expected to be highly sensitive to dosage variation changes in functionally relevant genes.

Our intent is to analyze orthologs of SLE GWA genes with confirmed risk for SLE with known connections to the TLR/IFN1 pathways or others that are tangential to these pathways. They will include *Blk*, *Irf5*, *Irf7*, *Irak1*, *Lyn*, *Stat4*, *Pcd1*, *Ptpn22*, and *Tnfsf4*, independently and in combinations. We will first determine dosage sensitivity effects when mice are hemizygous for single GWA genes. The hemizygous mice will be created by crosses of -/- KO ♀ mice to B6. *Yaa* ♂ mice. The KO mice to be used are bred at least 10 generations onto the C57BL/6 (B6) background. Sets of 16 age-matched ♂ mice of each hemizygous genotype will be produced and paired with an equal number of standard B6. *Yaa* mice. In our extensive experience, this number of biological replicate B6. *Yaa* mice will provide more than adequate power for the multiple, readily quantified phenotypes to be measured [48, 49].

We will then determine whether there is evidence for CHI by the genetic weakening of genes in permutational combinations. We seek to identify such gene combinations because they serve a dual purpose: 1) identify functionally relevant pathway interactions; and 2) suggest new genetic interactions that can be reevaluated in human SLE GWAS data. Our experimental design is to produce *Yaa* mice carrying hemizygous KO alleles in combination. These mice will be produced by crossing KO #1 X KO #2 F₁ ♀ mice to B6. *Yaa* ♂, and selecting doubly hemizygous ♂ progeny by KO-specific genotyping.

To assess alterations that emerge at an early age, half of the mice will be sacrificed at 8 weeks of age and the remaining 8 mice will be sacrificed at 20 weeks to assess more advanced autoimmune processes. Spleen samples will be prepared for flow cytometry. Multiparameter flow cytometry will be used to quantify the relative proportions and total numbers of splenic T cell, B cell subsets and myeloid cell subsets, including plasmacytoid dendritic cells. When possible, we will adhere to the antibody reagents and gating parameters on the *ImmGen* website (<http://immgen.org>). Included will be the analysis of B cells for activation markers/co-stimulatory molecules. Sera from the mice will be used to quantify the concentrations of IgM, IgG1, IgG2a, IgG2b and IgG3. All of these procedures are well-established in the Roopenian Laboratory [48, 50].

We will generate transcriptional data from early stages of autoimmune pathogenesis. These samples will come from the above-described 8 week old mice because the primary effects are less likely to be obscured by secondary changes at that age. We will focus on two key cellular components. The first is follicular B cells, which are the most differentiated B cell population prior to activation in the lymphoid germinal centers. The second is CD4⁺ T cells which are critical drivers of B cells and the pathogenic autoantibodies that they produce. Samples will be acquired through FACS sorting of cells preserved from 4 replicate samples (each a pool of 2 mice) per condition. We will perform preliminary experiments to guide our choice of gene expression technology, Affymetrix microarrays (1.0 ST MuGene Mouse 430) or RNAseq, that will be used to compare the consistency of the data and the depth of useful information that emanates. Once decided, data emanating from the chosen platform will be used for molecular pathway analyses described in Aim 3.

Results and expected outcomes

We will use the data accrued from standard B6. *Yaa* mice as our baseline. We will then seek deviations caused by each hemizygous KO or CHI gene combination. Given that the majority of genes tested are considered to be positive regulators of immune processes, we expect that some of them in the hemizygous state will cause phenotypic reductions in disease. However, *Tnaip3*, a negative regulator of immune processes, is predicted to show increases in disease when in the hemizygous state. Direct effects are expected to be evident from the subset and activation analysis of the B cell and CD4⁺ T cell compartments because the KO genes included in this study are known to impact those cell types. Subdivisions should be evident as well. Changes in the numbers of naïve B cells would suggest primary effects at the newly-formed B cell stage, while changes at late stages of maturation (the germinal center and plasmablast stages) and accompanied by class switched serum IgG would be consistent effects on autoimmune promotion. Changes in the proportions and activation state of T cells, especially, ICOS⁺ follicular CD4 T cells, will inform on whether the mutations impact these critical helper cells. Alterations in proportions and activation state determined by the upregulation of MHC Class II expression of myeloid-derived cells (monocytes, conventional DC and macrophages) or lymphocyte-derived pDC should also become evident. The overall results from this study will: i) indicate whether orthologs of SLE-risk genes exhibit dosage sensitivities that impact functionally relevant autoimmune processes; ii) provide

cellular and serological descriptions of the processes affected; and iii) by inference suggest additional GWA genes for testing.

Potential pitfalls and solutions

We do not foresee any complications in prosecuting the studies described because complex mouse crosses and biomarker studies with the *Yaa* lupus mouse model are major strengths of our laboratory. However, we realize that a potential concern is that most of the KO alleles to be analyzed are congenic and therefore carry genes of 129 embryonal stem cell parentage. The KO genes chosen do not lie in regions of significant disease risk in the *Yaa* model and we do not expect that 129 loci normally documented to have effect in the homozygous state will compromise our results. This concern can be abrogated in future studies by use of the newly emerging KO alleles on the B6 background made available by the KOMP initiative [51].

Aim 2b: Model reverse cholesterol transport with complex haploinsufficiency

Coronary artery disease (CAD) is the leading cause of death in the United States. Clinical and mouse studies show that a high plasma cholesterol level is a major predictor of CAD, whereas a high level of high-density lipoprotein (HDL) cholesterol protects against CAD [52]. The key role of HDL as a carrier of excess cellular cholesterol in the reverse cholesterol transport pathway (Figure 4) provides protection against atherosclerosis [53]. In fact, the most common lipid abnormality observed in patients coming to the emergency room with first heart attacks is low HDL.

Complex traits such as HDL cholesterol levels are the result of a combination of multiple genetic and environmental factors. Quantitative trait loci (QTL) analysis has been used to study the genetics of HDL levels in humans and mice. By 2005, approximately 40 HDL-regulating QTL had been identified in each species [53]. In the past two years, seven human genome-wide association (GWA) studies have confirmed known and identified new genomic loci associated with HDL levels [54-61].

To test complex haploinsufficiency for the reverse cholesterol transport (RCT) pathway in mice we will test a set of heterozygous knockout mice for their effect on liver expression (measured by RNAseq), HDL levels (measured in plasma), HDL function (measured by a cholesterol efflux assay), and atherosclerosis susceptibility (measured by histology of the aorta). We selected eight genes (*Apoa1*, *Apoa2*, *Apoa4*, *Apom*, *Abca1*, *Abcg1*, *Scarb1*, and *Ldlr*) for which knockouts on a C57BL/6J genetic background are available.

APOA1 is produced in the liver and excreted in the blood flow where it moves to peripheral tissues and cholesterol-loaded macrophages as lipid poor APOA1. It binds to ABCA1, which is a receptor that facilitates the transfer of cholesterol from the tissue or macrophage to APOA1. The particle, now known as pre- β HDL undergoes maturation (influenced by proteins like APOA4) and picks up other apolipoproteins (like APOA2 and APOM) and becomes an α -HDL particle, which is able to return to tissues and macrophages to pick up more cholesterol, this time using ABCG1 to facilitate the transfer. Finally the HDL particle returns to the liver where it binds to SRB1 (encoding by *Scarb1*) and, through associated APOE, to LDR, gets internalized, and APOA1 is recycled while the cholesterol is eliminated through the bile acid pathway.

For most of these lines effects on HDL levels have been described when in the heterozygous state. We will first characterize each single heterozygous knockout line to obtain a baseline. We then make double heterozygous knockout mice for each of the 28 possible combinations. For each individual knockout and each of the combinations we will generate 15 ♂ mice. Starting from homozygous knockout lines the breeding is

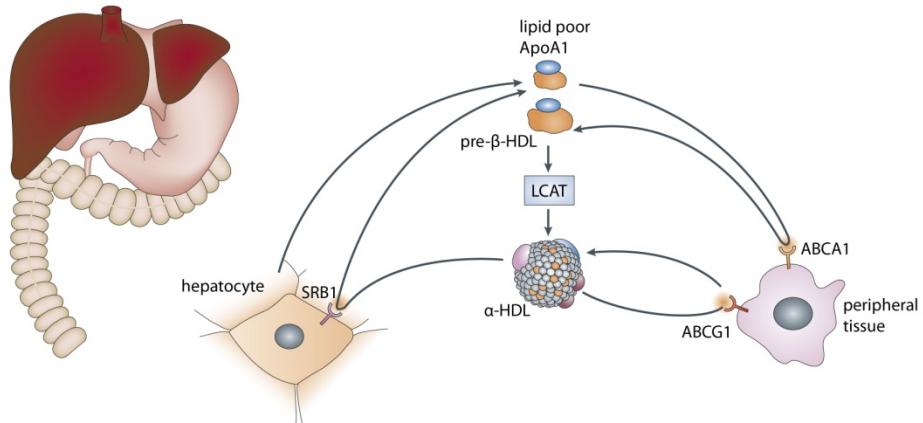


Figure 4. Schematic representation of the Reverse Cholesterol Transport pathway, which plays an important role in the elimination of excess cholesterol in the peripheral tissues and macrophages and thereby an important factor in the susceptibility for CAD.

simply mating the homozygous knockout lines with C57BL/6J mice in the case of the single knockouts, and with each other to obtain the double heterozygous F1 mice. As there is the possibility of a directional (epigenetic) effect we will breed each combination in both directions (e.g. we will generate 15 ♂s from a cross between an Apoa1^{+/-} ♀ and an Apoa2^{+/-} ♂ and 15 ♂s from a cross between an Apoa2^{+/-} ♀ and an Apoa1^{+/-} ♂). At 8 weeks of age all 15 mice of each group will move from a normal chow diet to an atherogenic diet. At 16 weeks, 5 animals will be sacrificed, blood will be collected for measuring HDL levels and cholesterol efflux capacity, and the liver will be collected for RNA isolation. The remaining 10 animals will be kept on the atherogenic diet for an additional 10 weeks. At 26 weeks of age the animals will be sacrificed and the heart and the aortic arch are collected. HDL levels are measured on a Beckman Coulter Synchron CX®5 Delta autoanalyzer. The cholesterol efflux is determined by differentiating human THP-1 monocytes to macrophages. Macrophage foam cell formation is induced by human oxidized LDL and fluorescent-labeled cholesterol. Cells are washed and incubated with serum from the mice to be tested. After removing medium, cells are lysed and homogenized. Fluorescence from the medium fraction and the cell lysate fraction are measured and the relative efflux is calculated. Liver RNA is used for RNAseq. The NEBNext mRNA Sample Prep Master Mix Set I kit will be used to prepare the sequencing libraries. These libraries will be sequenced single-end on an Illumina HiSeq 2000 instrument. Hearts are collected with the ascending aorta and fixed in 4% PFA overnight and embedded in

gelatin. Hearts are sectioned and sections are discarded until the end of the aortic sinus is found. Starting from the aortic sinus, 20 10µm sections are collected and stained with Oil Red O. Sections are scanned and the aorta is digitally reconstructed. After reconstruction, the volume of the Oil Red O stained lesions in the 200µm section is calculated.

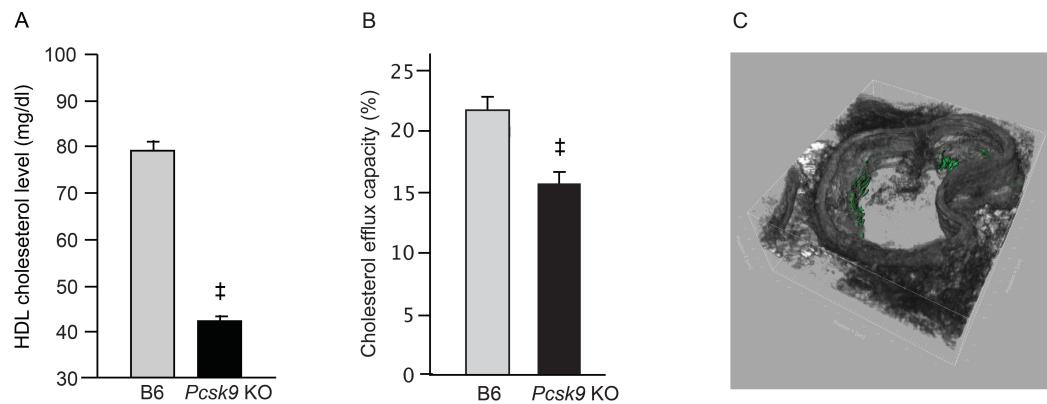


Figure 5. Comparison between male homozygous *Pcsk9* knockout and B6 wild-type animals for (A) HDL cholesterol levels, (B) cholesterol efflux capacity, and (C) atherosclerotic lesion size to demonstrate the major phenotypes measured in this study. *Pcsk9* knockout animals have lower HDL cholesterol levels caused by an absence of APOE in the HDL particles. As a result they are less efficient in taking up cholesterol from the macrophages. However, the reduced efficiency does not lead to an increase in atherosclerotic lesions after an 18-week atherogenic diet.

Results and expected outcomes

For the single heterozygous knockout lines we expect to identify sets of differentially expressed genes in the liver as a result of gene dosage of the target gene. For some lines these sets will overlap, while for other lines these will be unique. Combining the strains will show us whether for liver expression this is a simple addition of the differentially expressed genes (both in which genes are differentially expressed and the expression level) or whether the interaction leads to additionally affected genes (or even a return to normal levels). We will systematically compare the single and double hemizygote effects to infer how the genetic perturbations relate in a functional network (Aim 3). We will assess this network for novel and expected interactions. For example, we expect the receptor ABCG1 hemizygote to partially suppress the effects of ABCA1 hemizygosity, since ABCG1 is a downstream actor in the RCT pathway.

For several of the lines it is known that the heterozygous animals have a decreased HDL cholesterol level, but it is unknown whether this leads to a decrease in functionality (as measured by cholesterol efflux capacity) and results in increased atherosclerosis susceptibility. We expect that a decrease in gene dosage in some of the single heterozygotes will have an effect on cholesterol efflux, but that this will not be enough to lead to an increase in atherosclerosis susceptibility. However, we do expect that for several combinations the additive effect will cause an increase in atherosclerosis susceptibility where the individual knockouts did not.

Preliminary data for *Apoa4* in reciprocal F1 crosses between NOD and PWD mice, in which *Apoa4* expression was twofold higher in the (NODxPWD)F1 compared to the (PWDxNOD)F1. We found a difference in a transcription factor-binding site between PWD and NOD that likely leads to a difference in *Apoa4* expression.

When combined in an F1 animal it obviously matters which strain provides the paternal allele and which the maternal allele. By making reciprocal combinations of the double heterozygous knockouts we will be able to identify such epigenetic effects.

Potential pitfalls and solutions

All methods proposed in this sub aim are well established in our laboratory and have been used on homozygous knockout lines and ENU mutant lines. A disadvantage of these methods is that, until recently, they have only been used in small experiments comparing the mutant line with the wild-type control and that there is a relatively large variation in measuring HDL levels, cholesterol efflux, and atherosclerotic lesion size, which allows only picking up relatively large effects. The solution to this problem is a further standardization of the protocols and the use of high quality controls that are the same throughout the whole project. In the last few months we have been working on standardizing the cholesterol efflux assay and the testing of atherosclerotic lesions and making them suitable for larger scale projects (a manuscript for the new method for measuring atherosclerotic lesions is in preparation). We have been able to scale up the number of samples to be tested while improving quality and reducing variation in the data.

Specific Aim 3: Develop methods to analyze complex haploinsufficiency.

We will derive small-scale genetic interaction networks of interactions between the mouse orthologs of SLE and HDL GWA genes. These networks will represent a series of hypotheses of how mutations at multiple loci interact to influence phenotypes related to disease. This is a systematic method to place candidate genes in the fragile networks that underlie disease.

We will develop and test methods to infer network models of how individual hemizygous variants influence other variants and, in turn, downstream phenotypes. These phenotypes include cell type populations, surface markers, and gene expression for SLE (Aim 2a) and HDL cholesterol levels, cholesterol efflux, and liver gene expression for HDL (Aim 2b). Our approach will be focused on adapting a previously-developed method of modeling complex traits in yeast [62, 63]. This method models genetic interactions as quantitative influences using matrix decomposition, and is conceptually based on the classical genetic interaction approach of observing how genetic variants act and interact to reveal functional relationships such as activation, repression, and pathway order [64]. These influences are defined as positive or negative numbers that account for the fraction of a quantitative phenotype inferred to be caused by a variant's activity (Figure 6). The activity of each variant is simultaneously inferred as a model parameter that is influenced by other variant, and increases or decreases relative to a baseline assigned to the unperturbed gene. Each measured phenotype is modeled by multiple influences acting throughout the inferred network, and the use of multiple phenotypes resolves ambiguities in causal interpretations of interactions that often arise when single phenotypes are analyzed in isolation. Fitted model parameters will be assessed for statistical significance to define a network will of significant interactions. The result is a network model that provides the consistently best fit to multiple phenotypes measures across a panel of strains with genetic variation. We will also consider the alternative approach of rule-based modeling [65], which recasts phenotypes and parameters in discrete terms and aims to similarly infer network models by finding a consistent set of logical (Boolean) relationships between genes and phenotypes to fit the data.

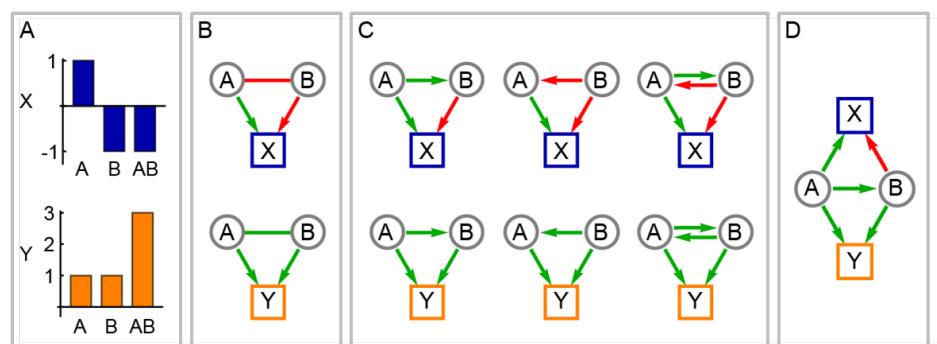


Figure 6. Modeling genetic interactions using multiple phenotypes. The complementary information in multiple phenotypes can be used to determine the directionality of interactions between genes with partially shared pleiotropy. **(A)** The effects of two hemizygous alleles, A and B, in isolation and combination on two quantitative phenotypes, X and Y. **(B)** Variants A and B show negative (red) epistasis for X and positive (green) epistasis for Y, in addition to their individual effects in terms of positive (green) and negative (red) directional edges. **(C)** Possible model interpretations for the negative and positive epistasis for X and Y, respectively. **(D)** The simplest model of influences between A and B that is consistent with the observations of both X and Y.

We recently developed this method by analyzing the effects of five yeast gene deletions on the mating pathway. The five deletions were randomly mixed in a population, similar to the research strategies we are considering in mouse models (Aims 2a and 2b). We independently measured two quantitative phenotypes: colony-level mating efficiency, and a molecular-level reporter of signaling pathway activity. We obtained a genetic interaction network (Figure 7) that correctly recapitulated the known genetics of the yeast mating pathway. This method is computationally fast and flexible, and can be straightforwardly applied to the mouse data we will collect.

Results and Expected Outcomes

The expected outcome of our analysis are models of how mouse orthologs of GWA genes interact to affect different aspects of the biology that underlies SLE and reverse cholesterol transport. In both cases, we expect to observe many additive effects, suggesting the genes operate in distinct pathways to contribute to disease. We also expect to find multiple genetic interactions between subsets of genes, forming cliques or near-cliques of densely interacting nodes in the network models. These network subsets will represent genes involved in the same pathway. By analyzing the patterns of activation and suppression, we will infer pathway ordering based on standard rules [64].

Potential Pitfalls and Solutions

In the course of developing our analysis methods we have identified a few potential problems when addressing a new data set. The method requires that the multiple phenotypes, to some degree, measure different aspects of the underlying biology and not be fully correlated. While this is highly unlikely to be a problem in the SLE analysis, which addresses multiple aspects of the immune system, it is a potential issue in the HDL transport analysis. In that case we will rely on a more supervised analysis of the gene expression data, in which we can identify expression patterns exhibited by sets of co-functional genes. Genes with different biological functions are expected to operate somewhat independently, which will be sufficient for our methods as has been demonstrated in our previous work [62, 63]. Alternatively, the data may encompass too much biological complexity for a single network analysis. The patterns of gene-to-gene influences may not be consistent across all phenotypes, and therefore yield no significant results. In this case, we will subdivide our phenotypes into those with the most biological similarity (e.g. a B cell phenotype combined with a gene expression pattern specific to B cells and a CD4⁺ Tcell phenotype combined with a gene expression pattern specific to CD4⁺ Tcells) and perform multiple analyses for each subset.

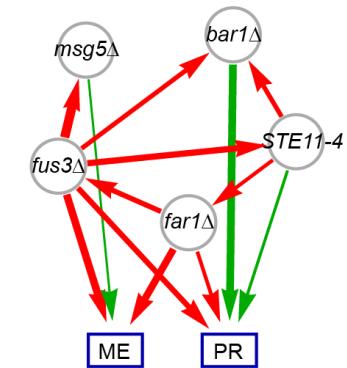


Figure 7. Network of genetic interactions in the yeast mating pathway. Complementary information in colony mating efficiency (ME) and molecular-level pheromone response (PR) was used to infer a network of interactions between gene perturbations. The method recovered the known genetic organization of the pathway, with mutations of downstream genes generally suppressing upstream mutations.

Specific Aim 4: Determine whether CHI patterns revealed are predictive of disease and its severity in a genetically diverse population.

Aims 2 and 3 will generate molecular network models of how perturbations of GWAS orthologs combine to affect SLE and CAD related phenotypes. These models, derived from comparing the effects of hemizygous genes individually and in pairs, will be networks of quantitative gene-to-gene and gene-to-phenotype influences with the capacity to make specific predictions for higher order combinations of mutations. Our goal is to predict the outcomes for these previously untested combinations and experimentally evaluate the predictions.

We will first make quantitative predictions of the SLE and CAD phenotypes we modeled in Aim 3. Since the networks represent mathematical equations we can simply adjust parameters from “wild-type” to “hemizygote” for each perturbed gene and compute the outcome [63]. The models will have been fit from single and double mutant combinations, so forward predictions can be made for any higher order combination (three hemizygotes, etc). We can also make predictions for the effects of gene over-expression [62]. Since all mathematical operations are simple matrix manipulations the predictions can be made in seconds. Forward prediction strategies such as this are the ultimate test of any mathematical model, and outcomes will critically depend on the network structures we derive in Aim 3.

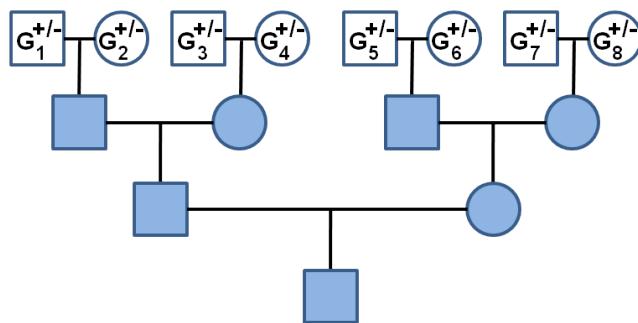


Figure 8. Outbreeding design to mix eight hemizygous genes (G_1-G_8). This design will be repeated 100 times and the zygosity of each gene will determined by PCR genotyping. For the SLE studies, the mice will be bred so that all males carry the *Yaa* sensitizer mutation.

genotypes, the large number of strains tested will allow us to identify the relative accuracy of each prediction. Single and double hemizygote combinations that result from outbreeding will be compared to the work in Aim 2 to assess the variance of each phenotype within and across different strains. The typical within-strain variance will be taken as the estimated experimental uncertainty for each strain. Prediction success will be gauged by the difference between predicted and experimental values relative to this estimated uncertainty, with complete success defined as prediction matching experiment within the estimated uncertainty.

Results and Expected Outcomes

Successful predictions will identify genetic subnetworks for which our model provides an accurate description. However, we expect many of our predictions to be incorrect. These instances of model failure will be especially informative. Since we have small-scale models of a few selected genes, we can hypothesize the most likely reason for model failure. For instance, predictions involving a certain gene might be especially inaccurate. This gene would then be prioritized for additional study by finding genes affected by its hemizygous allele in our data, and querying MGI to identify interacting proteins and coexpressed genes involved in the relevant biological system (SLE or CAD). The process can identify new genes for inclusion in a second round of experimentation and computational modeling. This iterative process of model refinement can be continued until convergence on an accurate model [62, 63].

Overall, this is a tractable strategy to build networks based in GWAS orthologs that model how these orthologs operate in combination within fragile, disease-causing pathways. This approach exploits the power of mouse models in understanding GWAS results [66] at a network level.

Potential Pitfalls and Solutions

The 100 outbred mice might fail to generate a random sampling of the 256 possible strains due to random chance or synthetic growth defects. To be certain we have sufficient diversity before the expense of phenotyping we will genotype all strains before proceeding. If some alleles are poorly represented we will breed and genotype additional mice. Once phenotyping is complete, the absence of replicate genotypes will potentially limit our certainty in novel biological discoveries. The genotypes that drive these discoveries will be reconstructed from our single-KO stocks and multiple individuals will be assayed for statistical confirmation of the novel results.

Beginning with the KO mice described in Aim 2, we will use an outbreeding design to generate arbitrary permutations of hemizygous alleles of up to 8 GWAS orthologs (Figure 8). A total of 256 unique genotypes will be possible and predictions will be made for all. However, acquiring all genotypes will be a logistical challenge and a random sample of 100 unique genotypes will be produced and analyzed for both SLE and CAD. These 100 mice will provide an unbiased source of complex genetic novelty to rigorously test our predictions of phenotypes and gene expression. Each mouse will be aged 8 weeks and subjected to the same quantitative disease-relevant phenotyping and GE assays as described in Aim 2. The data will be processed as in Aim 2 and the results will be directly compared to model predictions. Although the outcross design does not guarantee replication of

APPROPRIATENESS FOR THE TRANSFORMATIVE RESEARCH AWARDS INITIATIVE

This proposal outlines a new strategy to address the outstanding challenge of understanding how multiple candidate genes combine to affect disease using the most prevalent model organisms. It constitutes a powerful new paradigm for using mouse model systems to place GWAS findings in biological contexts, which will substantially improve the capacity to translate genetic analysis into clinical prognoses and treatments. It synthesizes advances in the understanding of evolved systems, network biology, and classical mouse genetics. It leverages the extensive knowledge, resources, and continuing investments in mouse genetics, and reinvigorates mouse research as a tool to understand human disease in the post-genomic age. It addresses the complexity of genetic systems, which often poses a challenge for studies based on single genes, by an experimental design based on the system-wide principle of biological robustness. Once developed through the studies outlined in this proposal, we foresee wide use of this approach to map the complex genetic etiology of many diseases.

TIMELINE

Year 1. Perform yeast CHI experiments (Aim 1) and assay hemizygous mice for haploinsufficiency (Aim 2). Collaboratively determine the optimal data processing to be followed in later SLE and CAD experiments.

Year 2. Analyze and publish yeast results (Aim 1). Analyze gene expression data for single hemizygotes from Year 1, assess if chosen genes are appropriate for CHI studies based on overlapping but partially distinct gene expression signatures. Select gene pairs for complex haploinsufficiency assays in mice (Aim 2) based on Year 1 results, and construct and assay up to 36 strains for SLE and 28 for CAD (Aim 2).

Year 3. Continue CHI studies in SLE and CAD (Aim 2) and analyze interactions to construct network models of SLE and CAD (Aim 3). If interaction analysis yields network models that fail to fit the data, reassess and identify additional genes to assess for CHI from prior pathway knowledge, gene expression results (Year 1), or current GWAS. Begin breeding hemizygous mice for complex outcross (Aim 4).

Year 4. Continue and analyze CHI data (Aim 3) and formulate predictions for all higher order combinations of hemizygous genes (Aim 4). Finish breeding and initiate assays of complex outcross mice (Aim 4).

Year 5. Finish assays of complex outcross populations and test predictions. Assess results, identify network interactions that lead to high prediction accuracy, and formulate hypotheses for failed predictions (Aim 4).

VERTEBRATE ANIMALS

1. Description of the proposed work. The goal of this study is to use laboratory mice as models to inquire into the genetic basis of two common human disorders, systemic lupus erythematosus (SLE) and coronary artery disease (CAD). The SLE project will main colonies of 9 selected knockout strains. The CAD project we will maintain colonies for 8 selected knockout strains. Control C57BL/6J and B6.Yaa strains will also be maintained. These knockout strains will first be used to characterize 15 and 16 male mice for each strain (approximately 250 mice). Subsequently we will use the strains to create 28 and 36 possible (reciprocal) double heterozygous knockout combinations and also characterize 15 to 16 male mice for each combination (approximately 1,000 mice). Finally, to generate an outcross population, we will mate mice from the double heterozygous groups for two generations in a manner to ensure equal chances of each knockout allele in the test population. We will generate two test populations of 100 mice for each disease. In all cases, samples will be obtained from euthanized mice for biological phenotyping and gene expression analyses.

2. Justification the use of animals. For studying complex haplotype insufficiency and its impact on SLE and CAD a mammalian animal model is required. The mouse is the model of choice because of the vast knowledge on genetics and genomics of this species and the ability to manipulate the genome. Every possible effort is made to gain the maximum amount of data or required biological materials from each mouse used in this study, so that there is no unnecessary sacrifice of life. We have implemented the Microsoft Access computer-based tracking system, Colony Management System (CMS), developed here at The Jackson Laboratory for all mouse breeding, genotypic, and inventory monitoring. The Mouse Tracking System permits an accurate updated inventory of all mouse strains, crosses, and inventory, thus minimizing unnecessary mouse breeding and genetic errors. All data generated are linked to the individualized mice within this database, thus maximizing the most efficient animal usage and data reliability.

3. Veterinary care of the animals. The Jackson Laboratory is accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC). Mice used in the project proposed in this application will be housed in the Research Animal Facility (RAF). No deviation from standard Jackson Laboratory protocols will be required. The Jackson Laboratory's Laboratory Animal Health Group, directed by Dr. Linda Waterman, a veterinarian board certified in Clinical Laboratory Animal Medicine, is responsible for veterinary care programs, operation of a diagnostic laboratory, and a quarantine facility for arriving animals or biological materials. Dr. Waterman manages the animal facilities. The veterinary clinical staff also includes one veterinarian board certified in microbiology, one in anatomic pathology, and a clinical veterinarian.

The health status of each animal room is evaluated every 13 weeks. This includes necropsy of representative clinically normal mice, culture of feces for pathogenic bacteria, fecal examinations for parasitic ova, serological screening for major mouse pathogens, and culture of selected organs for pathogenic bacteria. Preventive medicine programs also include microbiological monitoring of sterilizers, culture of environmental surfaces for coliforms, monitoring animal water for pH or residual chlorine, and culture of feed ingredients. Constant evaluation of skin lesions will be done to ensure that these mice do not develop severe or debilitating disease. Quantitative monitoring of disease severity is one of the criteria used for the gene mapping studies.

Animal care technicians are trained to identify mice that deviate from standards of normalcy for that strain. Deviant mice are brought to the attention of the PI and veterinary staff.

Four to five adult mice (depending upon body weight) or a single breeding pair with pre-weaned young are maintained in polycarbonate cages (50 square inches floor area). Mice are fed *ad libitum* with the NIH-31 6% fat mouse diet. Cages are washed once a week and filled with an appropriate amount of sterilized white pine shavings. Water bottles are also washed weekly and filled with either acidified (pH 2.8-3.2) or chlorinated water (100-15 ppm residual chlorine). Cage lids are covered with a non-woven polyester flat filter. Food hoppers, which are attached to the cage lids, are filled weekly. The water bottle on each cage is checked daily (Monday through Friday) for any sign of water delivery malfunction. All animals receive clean housing on a weekly basis. The light cycle is 12 hrs light:12 hrs dark.

4. Procedures for ensuring minimal discomfort, distress, pain, or injury. All procedures are reviewed and approved by our Institutional Animal Care and Use Committee. All personnel performing these procedures are or will be trained by the veterinary staff using IACUC approved methods. Blood is collected from the retro orbital sinus by an experienced technician. Tail tip amputation (< 6 mm) will be performed to obtain DNA for genotyping. Other biological samples are obtained after euthanasia. The overall experimental design minimizes the possibility that mice develop severe disease phenotypes. In the unexpected case of significant duress of the mice, the mice will be euthanized.

5. Methods of euthanasia. Mice will be euthanized by CO₂ asphyxiation (for adults) or CO₂ asphyxiation followed by decapitation (for neonatal mice). These methods of euthanasia are consistent with recommendations of the Panel of Euthanasia of the American Veterinary Medical Association.

MULTIPLE PD/PI LEADERSHIP PLAN

We (Drs. Roopenian, Carter, and Korstanje) have co-developed the proposed project and provide the highly complementary expertise that will insure its success. We are in agreement with our specific roles, which are readily discerned by our areas of expertise: Roopenian (autoimmune disease), Carter (computational biology), and Korstanje (cardiovascular disease).

While the participants in this project are expected to interact on an essentially daily basis, the PIs will meet more formally once a month to discuss planning aspects, new data, and to review outcomes and actions necessary. At this time, budget apportionments and expenditures will also be reviewed to make certain that they are on track. These meetings will include all personnel directly involved in the project to keep all on the same page and to reinforce the interface between the wet lab and computational aspects of the project.

Manuscripts, conference presentations, and reports will be prepared collaboratively and the PIs will share senior authorship when appropriate.

Carter will have final say on decisions related to biocomputational data, analysis and interpretation. Roopenian and Korstanje will have final say on decisions related to their spheres of *in vivo* experimentation, cellular and data analyses and interpretation. The Co-PIs each have a reputation of collegiality and an obvious enthusiasm to take advantage of this unusual collaborative opportunity. However, in the unlikely event that common agreement cannot be resolved by thoroughly discussing and carefully considering the pros and cons of specific actions, Dr. Robert Braun, our Director of Associate Director and Chair of Research, would be asked to intervene.

REFERENCES

1. Deng, Y. and B.P. Tsao, *Genetic susceptibility to systemic lupus erythematosus in the genomic era*. Nat Rev Rheumatol, 2010. **6**(12): p. 683-92.
2. Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits*. Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9362-7.
3. Manolio, T.A., et al., *Finding the missing heritability of complex diseases*. Nature, 2009. **461**(7265): p. 747-53.
4. Zuk, O., et al., *The mystery of missing heritability: Genetic interactions create phantom heritability*. Proceedings of the National Academy of Sciences of the United States of America, 2012.
5. Hartman, J.L.t., B. Garvik, and L. Hartwell, *Principles for the buffering of genetic variation*. Science, 2001. **291**(5506): p. 1001-4.
6. Kitano, H., *Biological robustness*. Nat Rev Genet, 2004. **5**(11): p. 826-37.
7. Rutherford, S.L., *From genotype to phenotype: buffering mechanisms and the storage of genetic information*. Bioessays, 2000. **22**(12): p. 1095-105.
8. Costanzo, M., et al., *The genetic landscape of a cell*. Science, 2010. **327**(5964): p. 425-31.
9. Surolia, I., et al., *Functionally defective germline variants of sialic acid acetylesterase in autoimmunity*. Nature, 2009. **466**(7303): p. 243-7.
10. Espinosa-Soto, C., O.C. Martin, and A. Wagner, *Phenotypic plasticity can facilitate adaptive evolution in gene regulatory circuits*. BMC Evol Biol, 2011. **11**: p. 5.
11. Nica, A.C., et al., *Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations*. PLoS Genet, 2010. **6**(4): p. e1000895.
12. Nicolae, D.L., et al., *Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS*. PLoS Genet, 2010. **6**(4): p. e1000888.
13. Weiss, L.A., et al., *A genome-wide linkage and association scan reveals novel loci for autism*. Nature, 2009. **461**(7265): p. 802-8.
14. Alkan, C., B.P. Coe, and E.E. Eichler, *Genome structural variation discovery and genotyping*. Nature reviews. Genetics, 2011. **12**(5): p. 363-76.
15. Gamazon, E.R., D.L. Nicolae, and N.J. Cox, *A Study of CNVs As Trait-Associated Polymorphisms and As Expression Quantitative Trait Loci*. PLoS Genet, 2011. **7**(2): p. e1001292.
16. Henrichsen, C.N., E. Chaignat, and A. Reymond, *Copy number variants, diseases and gene expression*. Hum Mol Genet, 2009. **18**(R1): p. R1-8.
17. Kondrashov, F.A. and E.V. Koonin, *A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications*. Trends Genet, 2004. **20**(7): p. 287-90.
18. Huang, N., et al., *Characterising and predicting haploinsufficiency in the human genome*. PLoS Genet, 2010. **6**(10): p. e1001154.
19. Haarer, B., et al., *Modeling complex genetic interactions in a simple eukaryotic genome: actin displays a rich spectrum of complex haploinsufficiencies*. Genes Dev, 2007. **21**(2): p. 148-59.
20. Hillenmeyer, M.E., et al., *The chemical genomic portrait of yeast: uncovering a phenotype for all genes*. Science, 2008. **320**(5874): p. 362-5.
21. Duncan, I.M., *Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster*. Genetics, 1982. **102**(1): p. 49-70.
22. Cutforth, T. and G.M. Rubin, *Mutations in Hsp83 and cdc37 impair signaling by the sevenless receptor tyrosine kinase in Drosophila*. Cell, 1994. **77**(7): p. 1027-36.
23. Karim, F.D., et al., *A screen for genes that function downstream of Ras1 during Drosophila eye development*. Genetics, 1996. **143**(1): p. 315-29.
24. Verheyen, E.M., et al., *Analysis of dominant enhancers and suppressors of activated Notch in Drosophila*. Genetics, 1996. **144**(3): p. 1127-41.
25. Kankel, M.W., D.M. Duncan, and I. Duncan, *A screen for genes that interact with the Drosophila pair-rule segmentation gene fushi tarazu*. Genetics, 2004. **168**(1): p. 161-80.
26. Artavanis-Tsakonas, S. and P. Simpson, *Choosing a cell fate: a view from the Notch locus*. Trends in genetics : TIG, 1991. **7**(11-12): p. 403-8.

27. McCright, B., et al., *Defects in development of the kidney, heart and eye vasculature in mice homozygous for a hypomorphic Notch2 mutation*. Development, 2001. **128**(4): p. 491-502.
28. McCright, B., J. Lozier, and T. Gridley, *Generation of new Notch2 mutant alleles*. Genesis, 2006. **44**(1): p. 29-33.
29. McDaniell, R., et al., *NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway*. American journal of human genetics, 2006. **79**(1): p. 169-73.
30. Warthen, D.M., et al., *Jagged1 (JAG1) mutations in Alagille syndrome: increasing the mutation detection rate*. Human mutation, 2006. **27**(5): p. 436-43.
31. Lozier, J., B. McCright, and T. Gridley, *Notch signaling regulates bile duct morphogenesis in mice*. PLoS One, 2008. **3**(3): p. e1851.
32. McCright, B., J. Lozier, and T. Gridley, *A mouse model of Alagille syndrome: Notch2 as a genetic modifier of Jag1 haploinsufficiency*. Development, 2002. **129**(4): p. 1075-82.
33. Tong, A.H. and C. Boone, *Synthetic genetic array analysis in Saccharomyces cerevisiae*. Methods in molecular biology, 2006. **313**: p. 171-92.
34. James, J.A., et al., *Systemic lupus erythematosus in adults is associated with previous Epstein-Barr virus exposure*. Arthritis Rheum, 2001. **44**(5): p. 1122-6.
35. Poole, B.D., et al., *Epstein-Barr virus and molecular mimicry in systemic lupus erythematosus*. Autoimmunity, 2006. **39**(1): p. 63-70.
36. Cooney, C.M., et al., *46,X,del(X)(q13) Turner's syndrome women with systemic lupus erythematosus in a pedigree multiplex for SLE*. Genes Immun, 2009. **10**(5): p. 478-81.
37. Tsokos, G.C., *Systemic lupus erythematosus*. N Engl J Med, 2011. **365**(22): p. 2110-21.
38. Rai, E. and E.K. Wakeland, *Genetic predisposition to autoimmunity--what have we learned?* Semin Immunol, 2011. **23**(2): p. 67-83.
39. Sanchez, E., et al., *Phenotypic associations of genetic susceptibility loci in systemic lupus erythematosus*. Ann Rheum Dis, 2011. **70**(10): p. 1752-7.
40. Scofield, R.H., et al., *Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus patients: support for the notion of a gene-dose effect from the X chromosome*. Arthritis Rheum, 2008. **58**(8): p. 2511-7.
41. Smith-Bouvier, D.L., et al., *A role for sex chromosome complement in the female bias in autoimmune disease*. J Exp Med, 2008. **205**(5): p. 1099-108.
42. Deane, J.A., et al., *Control of toll-like receptor 7 expression is essential to restrict autoimmunity and dendritic cell proliferation*. Immunity, 2007. **27**(5): p. 801-10.
43. Pisitkun, P., et al., *Autoreactive B cell responses to RNA-related antigens due to TLR7 gene duplication*. Science, 2006. **312**(5780): p. 1669-72.
44. Subramanian, S., et al., *A Tlr7 translocation accelerates systemic autoimmunity in murine lupus*. Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9970-5.
45. Richez, C., et al., *TLR4 ligands induce IFN-alpha production by mouse conventional dendritic cells and human monocytes after IFN-beta priming*. J Immunol, 2009. **182**(2): p. 820-8.
46. Flesher, D.L., et al., *Recent advances in the genetics of systemic lupus erythematosus*. Expert Rev Clin Immunol, 2010. **6**(3): p. 461-79.
47. Murphy, E.D. and J.B. Roths, *A Y chromosome associated factor in strain BXSB producing accelerated autoimmunity and lymphoproliferation*. Arthritis Rheum, 1979. **22**(11): p. 1188-94.
48. McPhee, C.G., et al., *MHC class I family proteins retard systemic lupus erythematosus autoimmunity and B cell lymphomagenesis*. J Immunol, 2011. **187**(9): p. 4695-704.
49. Kim, H.J., et al., *CD8+ T regulatory cells express the Ly49 Class I MHC receptor and are defective in autoimmune prone B6-Yaa mice*. Proc Natl Acad Sci U S A, 2011. **108**(5): p. 2010-5.
50. Bubier, J.A., et al., *A critical role for IL-21 receptor signaling in the pathogenesis of systemic lupus erythematosus in BXSB-Yaa mice*. Proc Natl Acad Sci U S A, 2009. **106**(5): p. 1518-23.
51. Guan, C., et al., *A review of current large-scale mouse knockout efforts*. Genesis, 2010. **48**(2): p. 73-85.
52. Gotto, A.M., Jr. and E.A. Brinton, *Assessing low levels of high-density lipoprotein cholesterol as a risk factor in coronary heart disease: a working group report and update*. Journal of the American College of Cardiology, 2004. **43**(5): p. 717-24.
53. Wang, X. and B. Paigen, *Genetics of variation in HDL cholesterol in humans and mice*. Circulation research, 2005. **96**(1): p. 27-42.
54. Aulchenko, Y.S., et al., *Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts*. Nature genetics, 2009. **41**(1): p. 47-55.

55. Chasman, D.I., et al., *Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis*. PLoS genetics, 2009. **5**(11): p. e1000730.
56. Kathiresan, S., et al., *A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study*. BMC medical genetics, 2007. **8 Suppl 1**: p. S17.
57. Kathiresan, S., et al., *Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans*. Nature genetics, 2008. **40**(2): p. 189-97.
58. Kathiresan, S., et al., *Common variants at 30 loci contribute to polygenic dyslipidemia*. Nature genetics, 2009. **41**(1): p. 56-65.
59. Sabatti, C., et al., *Genome-wide association analysis of metabolic traits in a birth cohort from a founder population*. Nature genetics, 2009. **41**(1): p. 35-46.
60. Saxena, R., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels*. Science, 2007. **316**(5829): p. 1331-6.
61. Willer, C.J., et al., *Newly identified loci that influence lipid concentrations and risk of coronary artery disease*. Nature genetics, 2008. **40**(2): p. 161-9.
62. Carter, G.W., et al., *Predicting the effects of copy-number variation in double and triple mutant combinations*. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 2012: p. 19-30.
63. Carter, G.W., et al., *Prediction of phenotype and gene expression for combinations of mutations*. Mol Syst Biol, 2007. **3**: p. 96.
64. Avery, L. and S. Wasserman, *Ordering gene function: the interpretation of epistasis in regulatory hierarchies*. Trends Genet, 1992. **8**(9): p. 312-6.
65. Shmulevich, I., E.R. Dougherty, and W. Zhang, *Gene perturbation and intervention in probabilistic Boolean networks*. Bioinformatics, 2002. **18**(10): p. 1319-31.
66. Capell, B.C., et al., *A farnesyltransferase inhibitor prevents both the onset and late progression of cardiovascular disease in a progeria mouse model*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(41): p. 15902-7.



Donnelly Centre for Cellular + Biomolecular Research



Brenda J. Andrews, Ph.D., FRSC
*Charles H. Best Chair of Medical Research
Professor and Chair, Banting & Best Department of Medical Research
Director, Terrence Donnelly Center for Cellular & Biomolecular Research
University of Toronto*

February 1, 2011

Gregory W. Carter, PhD
Assistant Professor
The Jackson Laboratory
207-288-6025

Dear Greg, Derry, and Ron,

I am writing to express my enthusiastic commitment to collaborate on projects described in your proposal entitled "Modeling Genetically Complex Human Diseases by Haploinsufficiency in Mice". Your plan to use networks of complex haploinsufficiency as a model for human disease promises to be a powerful means to understand how trait genes identified through genome-wide association studies interact to produce disease phenotypes. My lab can contribute substantially to this project by providing the tools, reagents and expertise for producing proof-of principle maps of complex haploinsufficiency using high-throughput studies of genetic interactions in hemizygous yeast.

This project is an exciting new application for the synthetic genetic array (SGA) technology the Boone and Andrews labs have developed over the past decade [Costanzo et al (2010). *Science* **327**, 425-431]. The double-hemizygote experiments will be simple to design and perform using the SGA approach, and we can also readily explore more complex interactions such as gene over-expression in hemizygous strain backgrounds. We have an excellent team with extensive experience in all aspects of the proposed yeast work as well as a unique robotics laboratory for automated SGA screens - would be pleased to host these studies in Toronto. I will also be available to consult or collaborate on any additional issues involving the analysis of large-scale data sets.

I believe this is an outstanding plan to further use the power of SGA technology to understand the basic genetic principles that underlie complex genetic interactions in yeast, mice, and humans. I wholeheartedly support your proposal and look forward to a productive collaboration.

Yours sincerely,

Donnelly Centre for Cellular + Biomolecular Research | 160 College Street Room 230 Toronto Ontario M5S 3E1 Canada
T 416.978.8562 | F 416.946.8253 | brenda.andrews@utoronto.ca | www.thedonnellycentre.utoronto.ca

Brenda J. Andrews, Ph.D., FRSC
Charles H Best Chair in Medical Research
Professor and Chair
Banting & Best Department of Medical Research
Director, Terrence Donnelly Center for Cellular and Biomolecular Research



Gary Churchill
Professor

January 5, 2012

Ron Korstanje
Greg Carter
Derry Roopenian
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609

Dear Ron, Greg, and Derry

I will be delighted to serve as a consultant to your NIH Director's Transformative Research Award application entitled "Complex Haploinsufficiency as a Model of Genetic Disease". In this proposal, you use a smart combination of the resources, novel high-throughput methods and vast experience currently available in mouse genetics to investigate this process that is clearly important to human health.

I will be able to advise you using my 15 years of mapping and microarray analysis experience at several stages of the project. Specifically, I will provide advice relating to the data analysis of RNAseq and microarray data and the reconstruction of gene networks.

I would also like to confirm that the phenotypic and gene expression data from this project will be supported and disseminated in an integrated format with CGD data.

I look forward to working with you and wish you the best of luck with your proposal.

Yours sincerely,

A handwritten signature in blue ink, appearing to read "Gary Churchill".

Gary Churchill

207-288-6189
207-288-6847 Fax
Gary.Churchill@jax.org
www.jax.org
600 or 610 Main Street
Bar Harbor, ME 04609-1500 or 1526

RESEARCH TOOLS AND RESOURCES SHARING PLAN

Research tools and resources will be made available in full accordance with the NIH Grants Policy Statement and the Principles and Guidelines for Recipients of NIH Research Grants and Contracts. The Jackson Laboratory (Jackson) has a leadership role in the biomedical research community in ensuring that mouse strains are readily available and that any legal restrictions on the sharing of these important research tools are entirely minimal. Indeed, Jackson worked closely with NIH to develop the Principles and Guidelines for obtaining and disseminating biomedical research resources, and enthusiastically endorses the research values and principles set forth in these NIH policies. The PIs of this application have an established record of sharing unique resources generated with NIH support including engineered and mouse stocks, protocols and data with the academic research community without any license or research restrictions.

Mice: The applicants are willing to make strains of mice produced with funds from this grant available following publication. Strains for which little or no demand are anticipated will be maintained by the investigators and distributed upon request as long as breeding stock is maintained. If live breeding colonies are discontinued, the strain may be available as cryopreserved sperm or embryos. Strains for which there is moderate to heavy demand will be offered to Jackson's Repository or Production units for distribution.

Over the past 60 years The Jackson Laboratory has had more extensive experience working with issues regarding the distribution of mice than any other institution, and has consistently endeavored to protect the basic research community from restrictions that would in any way inhibit the full use of the mouse as a research tool. Our distribution policy requires that donor institutions agree that Jackson can distribute mice provided by the donor to academic and non-profit investigators without any licensure requirements. Indeed, Jackson has for many years successfully resisted donor institutions from insisting on "reach-through" to option rights, royalties or products, which reach-through claims inhibit the use of mice as a research tool and are now strongly discouraged by NIH's Principles and Guidelines. Jackson distribution policy protects academic institutions from any licensure requirement, but does provide that donor institutions seeking financial return for the mice may request licensing arrangements directed at commercial use or use by for-profit companies. These license agreements are negotiated with these companies by the technology transfer departments of the donor institutions. The distribution of mice, however, may also be subject to patent rights of third parties for technologies used to make the mice, and again in these circumstances, Jackson has protected the basic research community from any licensing or license requirements that inhibit the use and availability of these important research tools.

Data and Software: The wealth of phenotypic and genotypic data emerging from the studies supported by this grant will be made available electronically following publication in accordance with N.I.H. policies. Knowing that errors in mouse pedigree and genotype propagate and compromise the quality of results that follow, we strictly adhere to the use of the Jackson Colony Management Software System. All breeders and experimental mice used by the Roopenian and Korstanje Laboratories are uniquely identified and tracked by this system and all pertinent data are linked to each experiment, mice and their pedigree. All gene expression data will be provided in a MAIME compliant format to the NCBI Gene Expression Omnibus public repository at time of publication, and the immunological gene expression data will also be offered to ImmGen. We will use the NIGMS-funded Center for Genome Dynamics (CGD) infrastructure currently existing at Jackson Lab to distribute data (<http://cgd.jax.org/datasets/datasets.shtml>). Computational and analytical tools will also be made available on the CGD portal (<http://cgd.jax.org/tools/tools.shtml>) and, when coded in R, via the Bioconductor online archive (<http://www.bioconductor.org>) as packages. Where appropriate, detailed protocols and data will be submitted to the Mouse Phenome Database, which is maintained at Jackson and curates data primarily on but not limited to inbred strains. Also where appropriate, data will be contributed to Mouse Genome Informatics, also housed at Jackson.

PHS 398 Checklist

OMB Number: 0925-0001

1. Application Type:

From SF 424 (R&R) Cover Page. The responses provided on the R&R cover page are repeated here for your reference, as you answer the questions that are specific to the PHS398.

* Type of Application:

New Resubmission Renewal Continuation Revision

Federal Identifier:

2. Change of Investigator / Change of Institution Questions

Change of principal investigator / program director

Name of former principal investigator / program director:

Prefix:

* First Name:

Middle Name:

* Last Name:

Suffix:

Change of Grantee Institution

* Name of former institution:

3. Inventions and Patents (For renewal applications only)

* Inventions and Patents: Yes No

If the answer is "Yes" then please answer the following:

* Previously Reported: Yes No

4. * Program Income

Is program income anticipated during the periods for which the grant support is requested?

Yes No

If you checked "yes" above (indicating that program income is anticipated), then use the format below to reflect the amount and source(s). Otherwise, leave this section blank.

*Budget Period	*Anticipated Amount (\$)	*Source(s)
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>

5. * Disclosure Permission Statement

If this application does not result in an award, is the Government permitted to disclose the title of your proposed project, and the name, address, telephone number and e-mail address of the official signing for the applicant organization, to organizations that may be interested in contacting you for further information (e.g., possible collaborations, investment)?

Yes No