***1. SPECIFIC AIMS***

We propose a novel approach to model the functional genomics of human systemic lupus erythematsosus (SLE) in laboratory mice. Like most autoimmune syndromes, the genetic predisposition to SLE is complex. This complexity is documented primarily by selective or genome-wide association studies (GWAS) in which allelic variation reveal individual genes with small, but significant, contributions to overall risk, but in concert result in a large phenotypic effect – SLE. While such information can be trangulated to knowledge-based genomic and protein databases to infer the potential molecular pathways affected, there is great need for empirical approaches that test such inferred pathways *in vivo* in the context of SLE-like autoimmune disease. We propose a novel application of the mouse model to elucidate the functional genomics of human SLE. The fact that genetic risk factors for SLE have small individual effects suggests these variants are not necessarily complete loss or gain of function alleles. Thus we parcel SLE genes into two broad groups of dosage variation: those with weakened function (hypomorphs) and those with strengthened function (hypermorphs). Motivated by the idea that the pathways disregulated by genetic variation are the pathways that are especially sensitive to changes in gene dosage, we have designed an experimental system to investigate the effects of gene dosage variation in a mouse model of SLE.

**Aim 1. Determine the impact of gene dosage changes of orthologs of human SLE-associated genes on mouse SLE-like disease.** We will focus on 7 genes that intersect with the key transcriptional controller, Interferon Response Factor 5, IRF5. We will reduce the copy numbers of these genes in mice genetically sensitized for SLE-like autoimmunity by the *Yaa* mutation (1, 2), and define the impact of under-dosed genes individually and in permutations on primary autoimmune processes.

**Aim 2. Identify the molecular pathways altered by gene dosage.** We will take a systems-based approach to relate cellular phenotypes and gene expression changes in B and myeloid cells caused by gene dosage variations. We will use novel computational algorithms to group these genes into pathways with specific functions.

Finally, the information from both Aims we be applied in an iterative manner, both to refine the pathways operative in mice and to establish correlates with human SLE. Our overall hypothesis is that the key regulatory pathways that determine susceptibility or resistance to SLE are those that are most sensitive to gene dosage.

***2. BACKGROUND***

**The nature of genetic variation in humans and its role in disease susceptibilities.**

Intraspecific genetic variation is dynamic and takes many forms. Sequence variants including single-nucleotide polymorphisms (SNPs), small insertions and deletions, and larger changes, such as gene copy number variation, all contribute to the diversity of phenotypic manifestations that characterize the human species and its diseases. SLE is one unfortunate manifestation. It is a prototypic complex genetic syndrome in which many genes (currently >30), many of which are shared with other autoimmune disorders, contribute to its genetic etiology and presumed environmental stimuli act as triggers (3-9). This complexity presents enormous challenges to understanding such disorders from a functional genomics perspective. However, even given this complexity, certain approximations can be made.

First, the contribution of each variant gene to overall risk is small. Variants identified by GWAS are based on simple sequence polymorphism (SNP) haplotyping and commonly contribute less than two-fold to overall risk. While in an early stage, targeted high-throughput genomic sequencing approaches designed to more accurately identify “rare” variants will potentially account for more of the risk (10). However, it still remains clear that each gene will only contribute incrementally to the complex genetics of SLE susceptibility. Thus, rather than complete loss or gain of protein function, the polymorphisms causing risk for SLE are better viewed as hypomorphic alleles that weaken or or hypermorphic alleles that strengthen their proteins’ functions to varying degrees. Second, in order for such alleles to contribute to pathological processes, they must function in finely-balanced, regulatory pathways. Finally, the triangulation of multiple risk-associated genes to the same regulatory pathway hallmarks the critical nature of that pathway in controlling the disease under study. These pathways can be considered to be “unbuffered” in that they are sensitive to small changes. Taken together, these approximations provide a conceptual framework for our novel application of the mouse model to extract biological meaning and clinical relevance from the human genetics of SLE.

**Conventional approaches to model SLE in mice**. The laboratory mouse will continue to be the primary 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 SLE-like autoimmunity have lead to invaluable insights. However, most focus on the genetic variation of mice and not humans.

Gene-specific approaches using conventional gene knockouts and transgenics have been similarly invaluable, but do not model the more subtle forms of allelic variation that confer risk for SLE in humans. Recapitulating human allelic variants in mice by genetic knock-in or transgenic technologies is generally regarded to be a promising bridge to model human SLE in mice. However, this forward thinking approach is very labor and resource intensive, fraught with uncertainties, and can be practically applied to only a fraction of candidate SLE-associated genes and polymorphisms.

**Our novel approach is based on the following conceptual and logistical framework:**

* The hypomorphic and hypermorphic genetic states evident in human SLE can be approximated by configuring orthologous mouse genes into hemizygous and homozygous states, respectively, by varying gene dosages.
* The impact of such dosage imbalances on the molecular and cellular processes characteristic of SLE-like autoimmune disease can be reliably measured.
* The data generated can be incorporated into biocomputational models, grounded by publicly available bioinformatic databases, to identify and remodel the pathways affected.
* The process is iterative. Genes whose involvement is supported by human SLE studies serve as the basis of experimental designs in mice. Gene interactions are then investigated by mouse zygosity studies. Novel gene interactions are then used to inquire the SLEGEN database for evidence of their genetic associations in humans, which, in turn, are used to form hypothesis that can be tested in mice.

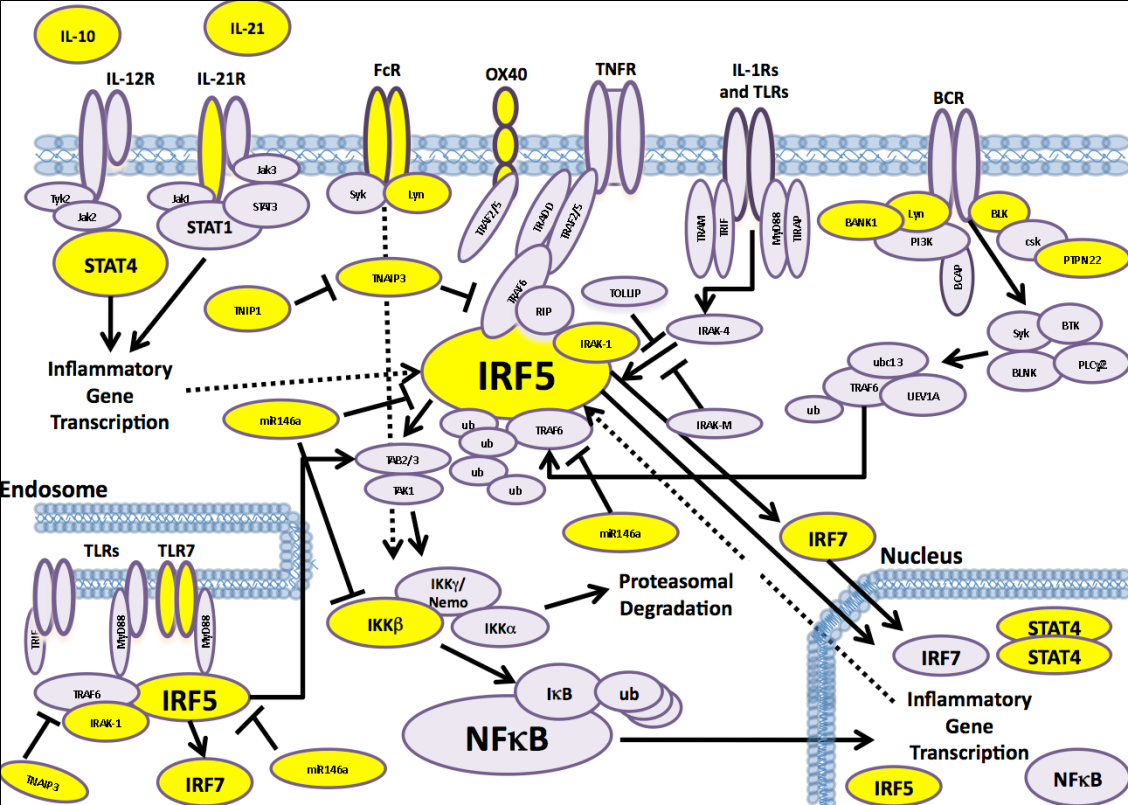
**Reasons to believe in the fertility of our approach.** Our proposal that gene dosage can be used to define the key regulatory pathways that break down in SLE is strongly grounded by the genetic principles of haploinsufficiency and the unfortunately termed phenomenon of “non-allelic non-complementation” (11). The Notch/Jagged signaling pathway is a particularly instructive example. The exquisite sensitivity of this pathway to haploinsufficiency is documented in Drosophila (12), mice (13, 14), and in humans with Alagille syndrome (15, 16). 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 (17, 18). With these principles in mind, as documented in multiple eukaryotic model organisms (19, 20) and humans (21), we note three important points. 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 non-allelic non-complementation. Third, dosage sensitivity has proven to be a powerful screen to identify interacting components of a regulatory pathway (12, 19).

We propose that these same principles can be applied to decipher the key regulatory pathways that give rise to SLE in humans. As is most easily documented by the X-chromosome, there are multiple lines of evidence to support dosage sensitivity in both human SLE and in mouse models. XXY men with Klinefelter's syndrome (a duplication of their X chromosome) have increased risk of SLE and women with Turner’s syndrome or an X chromosome-deletion have decreased risk (22, 23). Female sex-reversed (X0) mice show increased susceptibility to pristane-induced lupus (24). Finally, the *Yaa* mutation is a potent sensitizer for SLE-like autoimmunity, caused primarily by the genetic duplication of *Tlr7* as the result of a transposition (25-27), and is counteracted by haploinsufficiency in *Irf5*  (28).

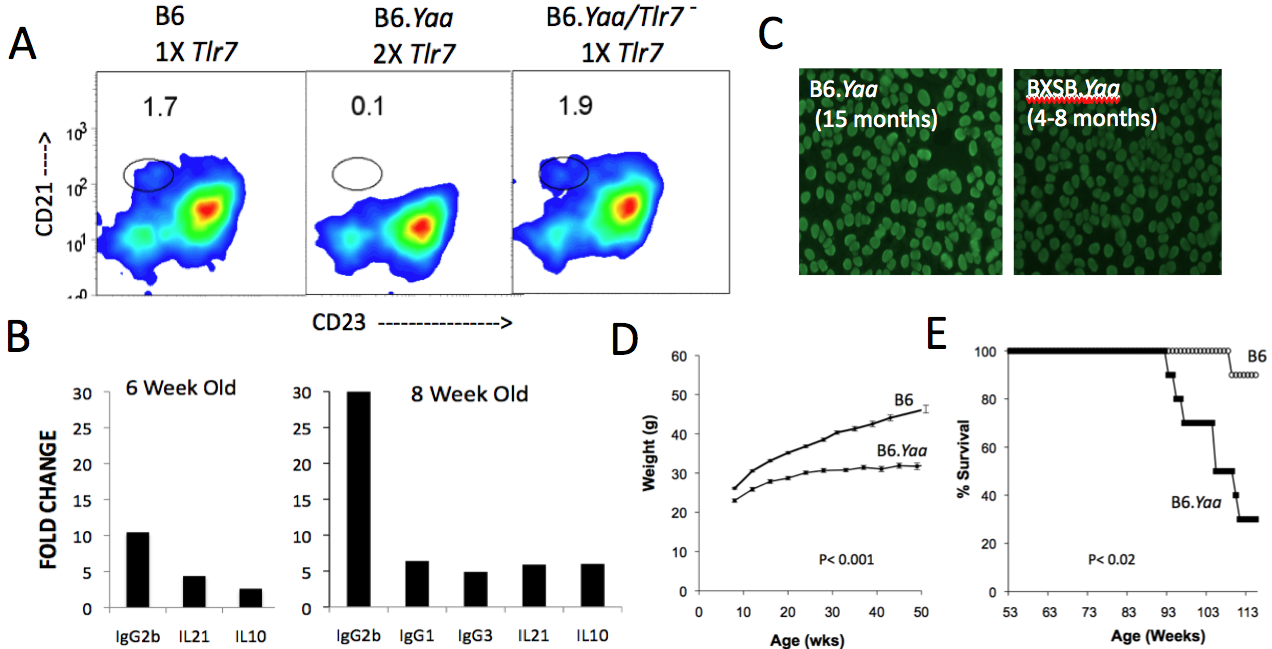
***3. PRELIMINARY STUDIES***

We propose to focus on the dosage sensitivity of SLE-risk associated genes that may influence the IRF5 regulatory pathway. As illustrated in **Figure 1**, kindly provided by our collaborators John Harley and Leah Kottyan, genetic association studies model IRF5 as a key central node. The importance of this pathway in SLE and related autoimmune disorders is extremely well-documented by functional studies in the immune regulation of B cells and dendritic cells (DC), monocytes and macrophages (28-31) (32) (33) (34). Importantly, IRF5 is also an essential component of the Type I IFN and TLR7/8 responses that themselves are genetically and functionally linked to SLE pathogenesis (28, 35, 36). However, such pathway models are not “real”, but instead serve to formulate hypotheses. The novel application of the mouse model that we propose provides a platform to empirically test such hypotheses *in vivo*.

The genetic background for the studies will be standard C57BL/6 (B6) for the following reasons: 1) on a B6 background, *Yaa* mice demonstrate easily measured primary abnormalities even at an early age (25, 37, 38) that predictably lead to a late onset form of SLE, with pathological manifestations similar to those observed in the more acute onset BXSB.*Yaa* model (examples, **Fig. 2A-E**, next page); 2) by using *Yaa* as a genetic sensitizer on the B6 background, we expect to be able to accurately and reliably evaluate gene dosage effects that either reduce or enhance these abnormalities; 3) we can conveniently incorporate knock-out alleles of pertinent genes that are commonly carried on the B6 background onto these B6.*Yaa* stocks; and 4) all data we generate will be compatible with the public Immunological Genome Project (ImmGen) database (www.immgen.org), which houses a wealth of carefully controlled microarray-based gene expression (GE) data of all known lymphocyte and myeloid lineages and normal subsets based on standard fetal to 10 week old B6 ♂ mice.

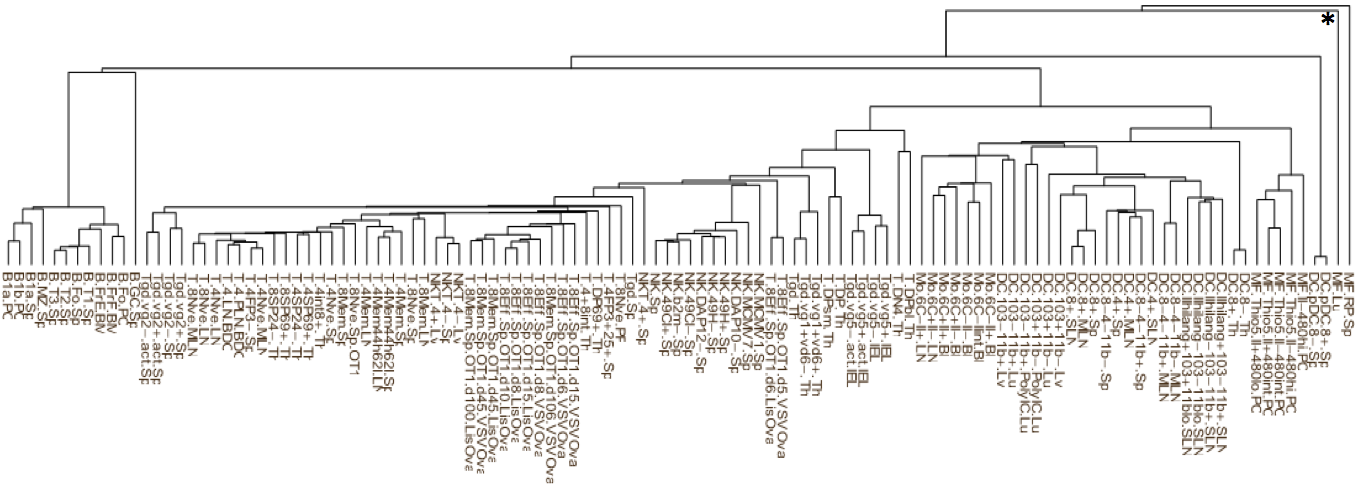


**Fig. 1. Model of the IRF5 regulatory pathway.** Genes in yellow are strong candidates at p<5x10E-8, and are generally replicated.



**Fig. 2. *Yaa* results in early onset aberrations and late onset pathological features on B6 background.** (**A**)B6.*Yaa* mice lack appreciable splenic MZ B cells (B220+CD21hiCd23lo) at 6 weeks of age, while B6 ♂ (a wild-type Y chromosome) and B6.*Yaa* ♂ mice reduced to have only 1 copy of *Tlr7* (B6.*Yaa/Tlr7-*) have a normal MZ B cell compartment. (**B**) *Yaa* causes the upregulation class switched IgGs along with the cytokines *Il21* and *Il10* at an early age. cDNA prepared from spleen cells from B6 and B6.*Yaa* ♂ mice were analyzed by qPCR. All fold change values shown are p ≤ 0.05 (n=3-5). (**C**) Anti-nuclear antibodies (ANA) similar to those arising at an earlier age in BXSB.*Yaa* mice develop in B6.*Yaa* at a later age. (**D**) *Yaa* results in less weight gain on a B6 background. *Yaa* shortens the lifespan of B6 mice (**E**), and results in late-onset renal disease (not shown).

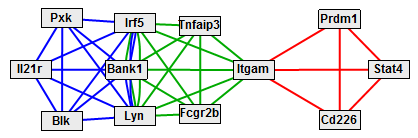
A major strength of our approach is that it is designed to take full advantage of the new and evolving ImmGen database ((39)(www.immgen.org)) to aid in preliminary hypotheses and interpret experimental outcomes. While human GE databases of immune cells are available, their utility is compromised by the variables inherent to sample acquisition and preparation, human genetic variability, and sporadic coverage of immune cell types. The extensive coverage of standardized immune cell categories in the ImmGen data provides a uniquely solid foundation for modeling immunological regulatory pathways, and the fact that both the ImmGen data and our studies based on B6 ♂ mice using the same DNA microarray platform and preparative conditions insures compatibility.

 Our co-investigator, Dr. Carter, has computationally evaluated the complete ImmGen GE dataset for its predictive value. More than 500 data files contributed by ImmGen were retrieved from Gene Expression Omnibus (GEO) database in CEL file format. These samples were co-normalized using the RMA algorithm (40) as implemented in *Bioconductor* software (41). Replicate samples were averaged for each probe and probes were mapped to genes and averaged to determine a single expression value of each gene. The expression of each gene in each sample was computed as a log2 ratio to progenitor hemopoietic bone marrow stem cells, so the baseline was this common originator cell population. The resulted in a matrix of log2-ratio expression values for 22,810 genes across 117 conditions. In order to evaluate a standard set of adult immune cells we did not include progenitor, precursor, fetal, or genetically altered cell samples.

**Fig. 3. Hierarchical dendrogram of 117 samples from Immgen.org.** From left to right, B cells, T cells (generally grouped into , , and activation), NK cells, and myeloid cells (subgrouped into monocytes, DCs, and macrophages). The outlier red pulp and lung macrophages are indicated by \*.

To assess how the gene expression patterns in the data relate to cell type and lineage relationships, we first performed hierarchical clustering on the samples (**Fig. 3**). Remarkably, standard cell subgroups and lineages were recapitulated from the gene expression data alone, with the only exceptions being red pulp and lung macrophages. These results demonstrate how GE patterns can be used to define the “steady state” status of discrete immune cell types and, by extrapolation, to identify deviations from this steady state as a result of genetic autoimmune processes.

We next examined the data in more detail to construct a broad map of gene expression patterns in the different cell types. For this analysis we used the method of singular value decomposition (SVD). SVD is a data-driven technique that finds and ranks the global patterns of gene expression, taking into account the full matrix of all genes by all conditions. Each expression pattern is defined by a subset of genes that exhibit differential expression across a subset of conditions. Finding stable patterns can be difficult due to non-deterministic properties of many algorithms (*e.g*. biclustering), but since SVD uses the machinery of matrix algebra there is a unique and entirely stable result for each data set. Using this method, we identified a number of interesting co-expression patterns. A common pattern was found that differentiated all mature cell types from the baseline stem cell sample. A second pattern identified a set of 440 genes up-regulated in myeloid cells and down-regulated in T cells. A third pattern identified a specific set of 261 genes up-regulated in B cells. These expression patterns provide a broad picture of how expression programs differ between cell types.

**Fig. 4. Network of SLE orthologs grouped by expression pattern.** Green edges connect genes up-regulated in myeloid cells, blue edges connect genes up-regulated in B cells, and red edges connect genes down-regulated in B cells. The 36 orthologs considered were: *Atg5, Atg7, Bank1, Blk, Cd226, Dnase1, Fcgr2b, Ica1, Infa1, Ifna2, Il10, Il21, Il21r, Irak1, Irf5, Itgam, Jazf1, Lyn, Mecp2, Myd88, Nmnat2, Prdm1, Pttg1, Pxk, Siae, Stat4, Stat5a, Stat6, Tnfaip3, Tnfsf4, Tnip1, Trex1, Tyk2, Ube2l3, Uhrf1bp1,* and *Xkr6*.

To understand how these patterns relate to the cellular disregulation that leads to SLE, we assembled a set of 36 mouse orthologs of genes discovered to be associated with SLE in human studies (**Fig. 4**). We inspected how each of the gene sets defined by the SVD analysis overlaps with these 36 SLE genes, in order to place each SLE gene in a global context of immune cell regulation (Fig. 4). We found that six of the SLE genes were up-regulated in myeloid cells: *Bank1*, *Fcgr2b*, *Irf5*, *Itgam*, *Lyn*, and *Tnfaip3*. Another six SLE genes were among those up-regulated in B cells: *Bank1*, *Blk*, *Il21r*, *Irf5*, *Lyn*, and *Pxk*. In both cases the gene sets contained more SLE genes than expected by random chance (*p* = 0.002 and *p* = 0.0003, respectively), suggesting that these expression patterns are associated with SLE-related biology. We note that this procedure only applies to SLE genes that are differentially regulated in normal cell lineages, and therefore additional genes listed in **Fig. 4** may also be associated with specific cell types. Nevertheless, this allows us to hypothesize that the genes *Bank1*, *Irf5*, and *Lyn* may affect SLE-related biology in both myeloid and B cell lineages, whereas the genes *Fcgr2b*, *Itgam*, and *Tnfaip3* are involved in myeloid dysregulation in SLE and *Blk*, *Il21r*, and *Pxk* are most likely involved in SLE-related B cell activity.

***4. EXPERIMENTAL DESIGN and METHODS***

**Aim 1. Determine the impact of gene dosage changes of orthologs of human SLE-associated genes on mouse SLE-like disease.**

**Overall Scheme.**

Our hypothesis is that the regulatory pathways that determine genetic risk for SLE are most sensitive to gene dosage effects. Given that dosage sensitivity has been documented for both *Tlr7* and *Irf5* in mouse SLE models (25-28), we propose that the IRF5 pathway and its intersection with TLR7 are highly sensitive. We will address this hypothesis by focusing on genes that are thought to impinge on the IRF5 pathway in B cells and myeloid cells. We will reduce gene dosage by creating hemizygous states of currently available KO alleles. To sensitize mice for SLE in a manner that will augment TLR7 signals, all mice will carry the *Yaa* mutation. To limit genetic heterogeneity, all studies will be performed on standard B6 background mice. Our experiments will study the primary cellular and serological abnormalities that the *Yaa* mutation is known to cause. We will first investigate whether there is evidence for dosage sensitivity of single candidate genes. Based on the outcome, we will then determine whether there is evidence for non-allelic non-complementation by permuting the zygosity of selected pairs of genes.

**Genetic design and phenotypic characterization.**

To determine dosage sensitivity effects when mice are hemizygous for single candidate genes, we will produce *Yaa* mice carrying the mutations indicated in **Table 1** below. The hemizygous mice will be created by crosses of -/- KO ♀mice to B6.*Yaa* ♂ mice. The KO mice to be used will have been bred at least 10 generations onto the B6 background. Sets of 14-15 age-matched ♂ mice of each hemizygous genotype will 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. We anticipate that 2 hemizygous groups and 1 standard B6.*Yaa* group can be practically managed in one experiment.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1. Mouse ortholog genes selected for dosage studies** | | | |
| KO Allele | Chr. | 1stReason | 2nd Reason |
| [*Irf5tm1Ttg*](http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=alleleDetail&key=37581) | 6 | SLE risk | Downstream of TLR7/8 and BCR |
| [*Irf7tm1Ttg*](http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=alleleDetail&key=37706) | 7 | “ | Downstream of IRF5/TLR7/8 |
| *Tnfaip3* | 10 | “ | Inhibition of IRF5 via TRAFf6 |
| [*Blktm1Tara*](http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=alleleDetail&key=8498) | 14 | “ | Downstream of BCR |
| [*Lyntm1Ard*](http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=alleleDetail&key=5382) | 4 | “ | Downstream of BCR/FCGRs |
| [*Irak1tm1Jth*](http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=alleleDetail&key=7963) | X | “ | Downstream of TLR7/8  IRF5 binding partner |
| *Il21rTm1wjl* | 7 | “ | Key B cell cytokine receptor |
| [*Stat4tm1Jni*](http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=alleleDetail&key=9022) | 1 | “ | Downstream of IL12R (and IL21R?) |

To assess alterations that emerge at an early age, half of the mice will be sacrificed at 8 weeks of age. The remaining mice will be sacrificed at 16 weeks to assess more advanced autoimmune processes.

Spleen samples will be prepared for flow cytometry. Multiparameter flow cytometry will be used to investigate the relative proportions and total numbers of splenic T cell, B cell and myeloid cell subsets. To insure that the results generated are based on and cross-comparable to standard B cell categorizations, whenever possible we will adhere to the antibody reagents and gating parameters outlined at the ImmGen website. Included will be the analysis of B cells for activation markers/co-stimulatory molecules using FAS and CD40 for B cells and ICOS and PD1 for CD4 T cells. Sera from the mice will be used to determine concentrations of IgM, IgG1,IgG2a, IgG2b and IgG3.

We will use the data accrued from standard B6.*Yaa* mouse as our baseline. We will then seek deviations caused by each hemizygous KO gene, starting with *Irf5*. Given that all genes with the exception of *Tnaip3* are positive regulators of immune processes, we expect that any of them that have an effect in the hemizgous state will show phenotypic bifurcations from *Yaa*. Direct effects are expected to be evident from the subset and activation analysis of the B cell and myeloid 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 B cells at the T1/T2 stage 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 with effects on autoimmune promotion. Alterations in the proportions of myeloid-derived cells (monocytes, conventional DC and macrophages) or lymphocyte-derived pDC should also become evident. Finally, 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. The overall results from this study will: 1) indicate whether orthologs of SLE-risk genes connected with the IRF5 regulatory pathway exhibit dosage sensitivities that impact functionally relevant autoimmune processes; and 2) provide cellular descriptions of the processes affected.

We will then we will then determine whether there is evidence for non-allelic non-complementation by the genetic weakening of genes in combination. We seek to identify such gene combinations because they serve a dual purpose: 1) clarify 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. We will select genes that did not result in dosage effects on their own but could be construed to fit into the same component pathway. For example, both *Irf7* and *Irak1* map the IRF5 pathway with connectivity to TLR7 signaling. Weakening of both of these genes may therefore result the attenuation of autoimmune processes while either alone would not do so. By extension, we can also use the approach to make connection that may not so apparent and therefore enrich pathway connections. For example, there are no current connections between *STAT4* and *IL21R*, but both engage JAKs and are SLE-risk loci. Finally, we it may be possible to identify counteracting effects on the same pathway. For example, *Tnfaip3* is considered to inhibit *Irf5* through its interaction with *Traf6*. Knowing that *Irf5* is dosage-sensitive (28), weakening of *Tnfaip3* by reducing its dosage to 1X may strengthen the IRF5 pathway to the extent that autoimmune processes are promoted.

**Considerations and opportunities.**

We do not foresee any complications in prosecuting the studies described becausecomplex mouse crosses and functional studies involving study of the *Yaa* mouse model are majorstrengths 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 significant disease risk in the *Yaa* model. Moreover, as the KO alleles will be in hemizygous configurations we do not expect that 129 loci normally documented to have effect in the homozygous state to compromise our results. Finally, the studies are forward looking. If substantiated, they can be extended to take advantage of the newly emerging KO alleles on a standard B6 background made available by the KOMP initiative (42).

**Aim 2. Identify the molecular pathways that are altered by gene dosage.**

**Overall Scheme.**

As alluded to in the Background, the computational integration of SLE-risk genes with the steady-state B6 ImmGen GE data reveals that specific subsets of SLE genes can be grouped into myeloid and/or B cell lineages. However, there is insufficient transcriptional information in the available steady state data to reliably model the regulatory pathways altered in a SLE setting. Here, we will use a systems-based approach to interrelate the changes caused by gene dose sensitivity at the transcriptional level to the cellular abnormalities that result from genetic sensitization by *Yaa*.

**Experimental design.**

Our objective is to generate transcriptional data from mice in early stages of autoimmune pathogenesis because the primary effects are less likely to be obscured by secondary changes at that stage. While the results in Aim 1 may revise our strategy, we provisionally plan to limit our analysis to two cell categories. The first is follicular B cells, which are the most differentiated B cell population prior to activation in the GC and critical components SLE-like disease. The second is pooled myeloid cells. Although the myeloid cells are heterogeneous (predominantly CD11b+ monocytes but also including conventional DCs, macrophages, etc.), we observed differential expression in our subset of SLE orthologs over most myeloid cell types. The approximation approach will provide a practical means to determine gene dosage effects that are common to or discern these cell types.

Samples will be acquired through FACS sorting of cells preserved from the same 8 week old mice analyzed in Aim 1. For compatibility with ImmGen data, our FACS will use antibody combinations, gating conditions, and RNA preparation conditions recommended by ImmGen (www.immgen.org/index\_content.html). Data from 4 replicate samples (each a pool of 2 mice) per condition will be generated on the Affymetrix platform using the Affymetrix 1.0 ST MuGene Mouse 430 29,000 gene chips. The raw data will be combined with the raw ImmGen data and co-normalized using R/Bioconductor (40, 43). Statistical significance of differential expression will be computed using R/maanova (44), with permutation analysis adjusted for multiple testing using the false discovery rate. Hierarchical clustering will be performed as previously (see Background) to verify that the overall expression profiles of the new B cell and myeloid samples cluster with the respective cell types in ImmGen.

To establish basic parameters, the two comparisons will first be made for cases of known dosage sensitivity: 1) B6 ♂ vs. B6.*Yaa* ♂ will define the transcriptional changes caused by *Yaa* (including its *Tlr7* duplication); 2) B6.*Yaa* ♂ vs. B6.*Yaa* *Irf5*+/-♂ will define the transcriptional changes effects of *Irf5* dosage. We then will determine the transcriptional profiles of the B6.*Yaa* ♂ mice hemizygous for remaining genes listed in **Table 1** (preceding page). We will combine the data from these mice along with those from B6.*Yaa* ♂, and B6.*Yaa* *Irf5*+/-♂ into a matrix of genes (rows) by strains (columns). All data will be normalized to B6 ♂ as a common baseline. We will perform singular value decomposition (SVD; see Background) on these data to identify the global patterns of gene expression (45, 46). From this, we expect to find: 1) an expression pattern common to all *Yaa* strains; 2) an expression pattern that aligns to the *Irf5* hemizygote and/or with other hemizygotes; and 3) potential pattern(s) that are unique to one or more of the other hemizygote genes and absent from *Irf5*, indicating activities independent of *Irf5*. We will identify genes that exhibit each expression pattern and query these gene sets for common function using the Gene Ontology database (47). These results will identify the biological processes affected by each hemizygous SLE ortholog.

To determine the transcriptional effects of paired hemizygous states, we will generate and add data from double-hemizygous B6.*Yaa* ♂ mice to our data matrix. Effects of the double-hemizygote that partially overlap with the single-hemizygote effects will be indicative of pathway cross-talk. By systematically analyzing how each double-hemizygote compares to the corresponding single-hemizygotes, we can construct a model of the pathways in which our SLE orthologs operate (48, 49). Parceling the expression data into follicular B cell and myeloid categories will add further discriminating information. For example, we might observe that *Irf5* and a few other hemizygotes (*e.g*. *Irak1*) similarly affect one function in myeloid cells, whereas in B cells *Irf5* may peripherally affect another function that is affected similarly by *Il21r*, *Blk*, and *Lyn*. This study will be based on classical pathway analysis as formalized by Avery and Wasserman (48) and applied to GE data with multiple pair-wise mutations (49).

We will also compare the observed patterns of gene expression to the cellular and serological phenotype data acquired in Aim 1. We expect to identify gene expression patterns that correlate with specific sets of phenotypes (*e.g*., hemizygous state in Gene X results in the reduction of FAS expression and a distinct transcriptional signature). Both simple and multi-dimensional correlates should emerge.

Finally, our GE data will be analyzed with the backdrop of the cell-type specific ImmGen GE data. We will use hierarchical cluster analysis to determine how well our GE data maps to ImmGen’s steady stage B cell and myeloid categories by comparing the lineage-clustered overall ImmGen database to our follicular B cell and the myeloid cell data. Such comparisons would reveal potentially altered cellular states that are sensitive to gene dosage.

**Considerations and opportunities.**

We recognize that the studies in this Aim are open-ended, designed to create rather than test hypotheses. However, the specific hypotheses generated ­– for example, any new candidate genes alone or in combination that are computationally inferred ­– can then be tested empirically for gene dosage effects *in vivo* in a convenient manner. We do not envision difficulties in generating high quality gene expression data because our Gene Expression Service has a consistent track record in delivering a high quality product. All gene expression data generated will be made available in a timely manner for public access through GEO and will adhere to community standards to enrich the Immgen database. Finally, Dr. Carter’s high level of computational and bioinformatics skills and the close interactions we expect to maintain with the Harley group insures connectivity between our studies in mice and investigation into human SLE.

***5. MILESTONES***

**Year 1.**

*Aim 1.* Completion of gene dosage studies *ex vivo and* biological monitoring and gene expression analysisof mice singly hemizygous for 4 knockout alleles and wild-type *Yaa* control groups.

*Aim 2.* Initial computational analysis of gene expression data and biological data and pathway inferences.

**Year 2.**

Aim 1. Completion of gene dosage studies *ex vivo* biological monitoring and gene expression analysisof mice singly hemizygous for 4 knockout alleles and ~3 doubly hemizygous knock-out allelic combinations. *Yaa* control groups.

Aim 2. High level and dimensional computational analyses, gene interactions, and evaluation of correlates in the SELGEN database by Harley group.

***6. SIGNIFICANCE***

We propose a novel application of the laboratory mouse to unravel the functional genomics of SLE. If validated, the same approach will have considerable impact in unraveling a broad range of other human genetic disorders. Information on variant genes that result in susceptibility for SLE is developing exponentially. This information clearly indicates that genetic susceptibility for SLE, as well as most other common genetic disorders, are highly complex. It is therefore a major challenge to translate these complex genetic patterns into a biological explanation of the aberrations that cause SLE, and even more of a challenge to apply this information in a manner that benefits SLE patients. Because of its remarkable genetic and experimental capabilities, the laboratory mouse in the only mammalian model that can be envisioned to full the large gap between complex genetics and the autoimmune processes that result. Our approach is expected to fill this gap in timely manner. The facts that risk factors for essentially all genetically complex diseases are characterized by small effects on genes that preferentially map to certain regulatory pathways (*e.g*., IRF5) is consistent with principles of gene dosage sensitivity that has proven to be an experimental tool for defining important finely tuned regulatory pathways. The broad range of knock-out alleles that are becoming increasingly available in mice provides the genetic tools to use dosage sensitivity as a means to query such regulatory pathways in a disease context. The quality of certain immunological bioinformatic databases adds special value systems based discovery of the pathways involved. Finally, by formulating the studies in mice to specifically address genes shown to impose risk in humans, the findings made have direct translational value for human SLE and related disorders.

***7. RELEVANCE* TO THE MISSION OF THE ALLIANCE FOR LUPUS RESEARCH**

Our proposal is designed to specifically address the goals outlined in ALR’s Functional Genomics and Molecular Pathways in Systemic Lupus Erythematosus initiative. Our approach is novel and exciting because it maximizes the potential of the mouse model to uncover the aberrant genetic and functional pathways that cause lupus in humans. It uses knowledge of human genes that are thought participate in human lupus to formulate functional genomic hypotheses that can only be reliably tested in a timely manner with mouse models. It fully capitalizes on important, current bioinformatics resources that are uniquely suited to the mouse model for systems based analyses. Finally, our approach employs advanced computational approaches to translate our discoveries in the context of human SLE.

***8. LITERATURE CITED***

1. Murphy, E. D., and J. B. Roths. 1979. A Y chromosome associated factor in strain BXSB producing accelerated autoimmunity and lymphoproliferation. *Arthritis Rheum* 22:1188-1194.

2. Andrews, B. S., R. A. Eisenberg, A. N. Theofilopoulos, S. Izui, C. B. Wilson, P. J. McConahey, E. D. Murphy, J. B. Roths, and F. J. Dixon. 1978. Spontaneous murine lupus-like syndromes. Clinical and immunopathological manifestations in several strains. *J Exp Med* 148:1198-1215.

3. James, J. A., B. R. Neas, K. L. Moser, T. Hall, G. R. Bruner, A. L. Sestak, and J. B. Harley. 2001. Systemic lupus erythematosus in adults is associated with previous Epstein-Barr virus exposure. *Arthritis Rheum* 44:1122-1126.

4. Poole, B. D., R. H. Scofield, J. B. Harley, and J. A. James. 2006. Epstein-Barr virus and molecular mimicry in systemic lupus erythematosus. *Autoimmunity* 39:63-70.

5. Rhodes, B., and T. J. Vyse. 2007. General aspects of the genetics of SLE. *Autoimmunity* 40:550-559.

6. Kelly, J. A., J. M. Kelley, K. M. Kaufman, J. Kilpatrick, G. R. Bruner, J. T. Merrill, J. A. James, S. G. Frank, E. Reams, E. E. Brown, A. W. Gibson, M. C. Marion, C. D. Langefeld, Q. Z. Li, D. R. Karp, E. K. Wakeland, M. Petri, R. Ramsey-Goldman, J. D. Reveille, L. M. Vila, G. S. Alarcon, R. P. Kimberly, J. B. Harley, and J. C. Edberg. 2008. Interferon regulatory factor-5 is genetically associated with systemic lupus erythematosus in African Americans. *Genes Immun* 9:187-194.

7. Tsuchiya, N., I. Ito, and A. Kawasaki. Association of IRF5, STAT4 and BLK with systemic lupus erythematosus and other rheumatic diseases. *Nihon Rinsho Meneki Gakkai Kaishi* 33:57-65.

8. Deng, Y., and B. P. Tsao. Genetic susceptibility to systemic lupus erythematosus in the genomic era. *Nat Rev Rheumatol* 6:683-692.

9. Harley, I. T., K. M. Kaufman, C. D. Langefeld, J. B. Harley, and J. A. Kelly. 2009. Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. *Nat Rev Genet* 10:285-290.

10. Surolia, I., S. P. Pirnie, V. Chellappa, K. N. Taylor, A. Cariappa, J. Moya, H. Liu, D. W. Bell, D. R. Driscoll, S. Diederichs, K. Haider, I. Netravali, S. Le, R. Elia, E. Dow, A. Lee, J. Freudenberg, P. L. De Jager, Y. Chretien, A. Varki, M. E. Macdonald, T. Gillis, T. W. Behrens, D. Bloch, D. Collier, J. Korzenik, D. K. Podolsky, D. Hafler, M. Murali, B. Sands, J. H. Stone, P. K. Gregersen, and S. Pillai. Functionally defective germline variants of sialic acid acetylesterase in autoimmunity. *Nature* 466:243-247.

11. Yook, K. J., S. R. Proulx, and E. M. Jorgensen. 2001. Rules of nonallelic noncomplementation at the synapse in Caenorhabditis elegans. *Genetics* 158:209-220.

12. Artavanis-Tsakonas, S., C. Delidakis, and R. G. Fehon. 1991. The Notch locus and the cell biology of neuroblast segregation. *Annu Rev Cell Biol* 7:427-452.

13. McCright, B., J. Lozier, and T. Gridley. 2006. Generation of new Notch2 mutant alleles. *Genesis* 44:29-33.

14. McCright, B., X. Gao, L. Shen, J. Lozier, Y. Lan, M. Maguire, D. Herzlinger, G. Weinmaster, R. Jiang, and T. Gridley. 2001. Defects in development of the kidney, heart and eye vasculature in mice homozygous for a hypomorphic Notch2 mutation. *Development* 128:491-502.

15. Warthen, D. M., E. C. Moore, B. M. Kamath, J. J. Morrissette, P. Sanchez, D. A. Piccoli, I. D. Krantz, and N. B. Spinner. 2006. Jagged1 (JAG1) mutations in Alagille syndrome: increasing the mutation detection rate. *Hum Mutat* 27:436-443.

16. McDaniell, R., D. M. Warthen, P. A. Sanchez-Lara, A. Pai, I. D. Krantz, D. A. Piccoli, and N. B. Spinner. 2006. NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *Am J Hum Genet* 79:169-173.

17. Lozier, J., B. McCright, and T. Gridley. 2008. Notch signaling regulates bile duct morphogenesis in mice. *PLoS One* 3:e1851.

18. McCright, B., J. Lozier, and T. Gridley. 2002. A mouse model of Alagille syndrome: Notch2 as a genetic modifier of Jag1 haploinsufficiency. *Development* 129:1075-1082.

19. Deutschbauer, A. M., D. F. Jaramillo, M. Proctor, J. Kumm, M. E. Hillenmeyer, R. W. Davis, C. Nislow, and G. Giaever. 2005. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169:1915-1925.

20. Delneri, D., D. C. Hoyle, K. Gkargkas, E. J. Cross, B. Rash, L. Zeef, H. S. Leong, H. M. Davey, A. Hayes, D. B. Kell, G. W. Griffith, and S. G. Oliver. 2008. Identification and characterization of high-flux-control genes of yeast through competition analyses in continuous cultures. *Nat Genet* 40:113-117.

21. Huang, N., I. Lee, E. M. Marcotte, and M. E. Hurles. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 6:e1001154.

22. Scofield, R. H., G. R. Bruner, B. Namjou, R. P. Kimberly, R. Ramsey-Goldman, M. Petri, J. D. Reveille, G. S. Alarcon, L. M. Vila, J. Reid, B. Harris, S. Li, J. A. Kelly, and J. B. Harley. 2008. 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* 58:2511-2517.

23. Cooney, C. M., G. R. Bruner, T. Aberle, B. Namjou-Khales, L. K. Myers, L. Feo, S. Li, A. D'Souza, A. Ramirez, J. B. Harley, and R. H. Scofield. 2009. 46,X,del(X)(q13) Turner's syndrome women with systemic lupus erythematosus in a pedigree multiplex for SLE. *Genes Immun* 10:478-481.

24. Smith-Bouvier, D. L., A. A. Divekar, M. Sasidhar, S. Du, S. K. Tiwari-Woodruff, J. K. King, A. P. Arnold, R. R. Singh, and R. R. Voskuhl. 2008. A role for sex chromosome complement in the female bias in autoimmune disease. *J Exp Med* 205:1099-1108.

25. Deane, J. A., P. Pisitkun, R. S. Barrett, L. Feigenbaum, T. Town, J. M. Ward, R. A. Flavell, and S. Bolland. 2007. Control of toll-like receptor 7 expression is essential to restrict autoimmunity and dendritic cell proliferation. *Immunity* 27:801-810.

26. Pisitkun, P., J. A. Deane, M. J. Difilippantonio, T. Tarasenko, A. B. Satterthwaite, and S. Bolland. 2006. Autoreactive B cell responses to RNA-related antigens due to TLR7 gene duplication. *Science* 312:1669-1672.

27. Subramanian, S., K. Tus, Q. Z. Li, A. Wang, X. H. Tian, J. Zhou, C. Liang, G. Bartov, L. D. McDaniel, X. J. Zhou, R. A. Schultz, and E. K. Wakeland. 2006. A Tlr7 translocation accelerates systemic autoimmunity in murine lupus. *Proc Natl Acad Sci U S A* 103:9970-9975.

28. Richez, C., K. Yasuda, R. G. Bonegio, A. A. Watkins, T. Aprahamian, P. Busto, R. J. Richards, C. L. Liu, R. Cheung, P. J. Utz, A. Marshak-Rothstein, and I. R. Rifkin. IFN regulatory factor 5 is required for disease development in the FcgammaRIIB-/-Yaa and FcgammaRIIB-/- mouse models of systemic lupus erythematosus. *J Immunol* 184:796-806.

29. Savitsky, D. A., H. Yanai, T. Tamura, T. Taniguchi, and K. Honda. Contribution of IRF5 in B cells to the development of murine SLE-like disease through its transcriptional control of the IgG2a locus. *Proc Natl Acad Sci U S A* 107:10154-10159.

30. Schoenemeyer, A., B. J. Barnes, M. E. Mancl, E. Latz, N. Goutagny, P. M. Pitha, K. A. Fitzgerald, and D. T. Golenbock. 2005. The interferon regulatory factor, IRF5, is a central mediator of toll-like receptor 7 signaling. *J Biol Chem* 280:17005-17012.

31. Yasuda, K., C. Richez, J. W. Maciaszek, N. Agrawal, S. Akira, A. Marshak-Rothstein, and I. R. Rifkin. 2007. Murine dendritic cell type I IFN production induced by human IgG-RNA immune complexes is IFN regulatory factor (IRF)5 and IRF7 dependent and is required for IL-6 production. *J Immunol* 178:6876-6885.

32. Krausgruber, T., D. Saliba, G. Ryzhakov, A. Lanfrancotti, K. Blazek, and I. A. Udalova. IRF5 is required for late-phase TNF secretion by human dendritic cells. *Blood* 115:4421-4430.

33. Mancl, M. E., G. Hu, N. Sangster-Guity, S. L. Olshalsky, K. Hoops, P. Fitzgerald-Bocarsly, P. M. Pitha, K. Pinder, and B. J. Barnes. 2005. Two discrete promoters regulate the alternatively spliced human interferon regulatory factor-5 isoforms. Multiple isoforms with distinct cell type-specific expression, localization, regulation, and function. *J Biol Chem* 280:21078-21090.

34. Couzinet, A., K. Tamura, H. M. Chen, K. Nishimura, Z. Wang, Y. Morishita, K. Takeda, H. Yagita, H. Yanai, T. Taniguchi, and T. Tamura. 2008. A cell-type-specific requirement for IFN regulatory factor 5 (IRF5) in Fas-induced apoptosis. *Proc Natl Acad Sci U S A* 105:2556-2561.

35. Kozyrev, S. V., and M. E. Alarcon-Riquelme. 2007. The genetics and biology of Irf5-mediated signaling in lupus. *Autoimmunity* 40:591-601.

36. Pascual, V., L. Farkas, and J. Banchereau. 2006. Systemic lupus erythematosus: all roads lead to type I interferons. *Curr Opin Immunol* 18:676-682.

37. Amano, H., E. Amano, T. Moll, D. Marinkovic, N. Ibnou-Zekri, E. Martinez-Soria, I. Semac, T. Wirth, L. Nitschke, and S. Izui. 2003. The Yaa mutation promoting murine lupus causes defective development of marginal zone B cells. *J Immunol* 170:2293-2301.

38. DesJardin, L. E., E. J. Butfiloski, E. S. Sobel, and J. Schiffenbauer. 1996. Hyperproliferation of BXSB B cells is linked to the Yaa allele. *Clin Immunol Immunopathol* 81:145-152.

39. Heng, T. S., and M. W. Painter. 2008. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* 9:1091-1094.

40. Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249-264.

41. Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.

42. Guan, C., C. Ye, X. Yang, and J. Gao. A review of current large-scale mouse knockout efforts. *Genesis* 48:73-85.

43. Irizarry, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31:e15.

44. Churchill, G. A. R/maanova

45. Alter, O., P. O. Brown, and D. Botstein. 2000. Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci U S A* 97:10101-10106.

46. Carter, G. W., S. Rupp, G. R. Fink, and T. Galitski. 2006. Disentangling information flow in the Ras-cAMP signaling network. *Genome Res* 16:520-526.

47. Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25-29.

48. Avery, L., and S. Wasserman. 1992. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet* 8:312-316.

49. Carter, G. W., S. Prinz, C. Neou, J. P. Shelby, B. Marzolf, V. Thorsson, and T. Galitski. 2007. Prediction of phenotype and gene expression for combinations of mutations. *Mol Syst Biol* 3:96.