Award Number: W81XWH-09-1-0391

TITLE: Host Genes and Resistance/Sensitivity to Military Priority Pathogens

PRINCIPAL INVESTIGATOR: Dr. Gerald I. Byrne

CONTRACTING ORGANIZATION: University of Tennessee
Memphis, TN 38103

REPORT DATE: June 2011

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-06-2011	Annual	1 JUN 2010 - 31 MAY 2011

4. TITLE AND SUBTITLE
Host Genes and Resistance/Sensitivity to Military Priority Pathogens

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-09-1-0391

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Gerald I. Byrne

E-Mail: omahdi@uthsc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Tennessee
Memphis, TN 38103

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The major objective of the award is to identify host genetic loci/pathways whose expression/lack of expression correlate with resistance vs. susceptibility to pathogenic challenge using recombinant inbred BXD mice. The pathogens studied are Francisella tularensis (FT), Burhholderia pseudomallei (Bp), Acinetobacter baumannii (Ab), Leishmania major (Lm), SARS, H5N1 avian influenza. A new FT reporter plasmid has been constructed resulting in bioluminescence production, which could be used to follow FT growth in vitro and in vivo. Preliminary observations on differential susceptibility to Bp 1026-mediated disease in BXD mouse strains have identified two suggestive QTLs, encoded on Chr 4 and 7. We are also carrying out parallel studies to identify cellular pathways responsible for genetic susceptibility to Bp infection using mice deficient in the inflammasome. We have identified 33 potentially useful phenotypic differences between B6 and D2 mice infected with Acinetobacter, and a QTL on Chr10 that may correlate with immune responsiveness. The Influenza project has identified a significant QTL associated with early production of proinflammatory cytokines on Chr 6. The Lm project has identified 2 suggestive loci regulating parasite burden. The mouse genomics core is continuing to develop new BXD strains. The bioinformatics core has begun analysis of data obtained from BXD mouse strains infected with Acinetobacter.

15. SUBJECT TERMS
Host genetics, Pathogens, Biodefense, Disease susceptibility, Humanized mice, Francisella tularensis, Burkholderia pseudomallei, Acinetobacter baumannii, Leishmania major, SARS, H5N1 avian influenza

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	70	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

This is our second annual progress report representing work done on the grant in the past year. We have included updates from previous reports as well as new data. We will relate our work progress to the milestones outlined in the grant proposal.

We hypothesized that the differential susceptibility to DoD-priority pathogens is the result of host genetic variability, and that these discreet loci and/or gene pathways can be identified using BXD recombinant inbred mice. The DoD priority pathogens are naturally occurring endemic diseases, emerging infectious diseases, and potential biowarfare agents: *Leishmania major*, *Burkholderia pseudomallei*, Severe Acute Respiratory Syndrome (SARS-CoV), highly pathogenic H5N1 Avian Influenza virus, *Francisella tularensis* and multidrug resistant *Acinetobacter baumannii*.

We outlined a three-step sequence for studying each pathogen with the following specific aims:

Objective 1 (18 months): To identify specific phenotypic differences in the responses of the BXD parental strains following infection with each of the DoD-priority pathogens.

Objective 2 (6 to 30 months): To identify host genetic loci and pathways that correlate with the differential susceptibility/resistance phenotype(s) of the parental mouse strains to the DoD-priority pathogens.

Objective 3 (12 to 36 months): To define and validate candidate genes and gene networks responsible for differential susceptibility/resistance phenotype(s) of the parental mouse strains for each DoD-priority pathogen.

Each project is performed in parallel to achieve the milestones set forth in Objectives 1-3. The infection models for each pathogen are at various levels of development, and therefore the timelines for the objectives differ according to the pathogen. This report covers the progress of work carried out to date.

The format of this report will present the information on the ongoing projects for the different pathogens, which includes descriptions of experimental procedures and results, and will highlight key research accomplishments and reportable outcomes. We will also describe plans for future work. Information from the animal and bioinformatics cores is also included in this report.

Key research accomplishments, reportable outcomes, conclusion, references

1. *Projects with DoD-priority Bacterial Pathogens*

1.1 Studies with *Francisella tularensis* (FT)

1.1.1 Generation of luminescence reporter for *Francisella tularensis*

A. Introduction

One of the objectives of the proposed research is to identify phenotypic differences in the responses of the BXD parental strains following infection with DoD-priority pathogens. To this end we have developed a *Francisella tularensis* shuttle vector that constitutively expresses the *Photorhabdus luminescens* lux operon in type A and type B strains of *F. tularensis*. The bioluminescence reporter plasmid was introduced into the live vaccine strain of *F. tularensis* and used to follow *F. tularensis* growth in a murine intranasal challenge model in real time by bioluminescence imaging. The results show that the new bioluminescence reporter plasmid represents a useful tool for tularemia research that is suitable for following *F. tularensis* growth in both in vitro and in vivo model systems. The use of this reporter in *F. tularensis* will facilitate the identification of phenotypic differences in challenge studies with the BXD mice.

B. Background.

Francisella tularensis is a gram-negative facultative intracellular bacterium that causes the zoonotic disease tularemia. *F. tularensis* infection of humans can occur by a number of routes, including the handling of infected animals, arthropod bites (Evans, 1985; Francis, 1937; Tarnvik, 1989), ingestion (Anda et al., 2001; Greco et al., 1987; Karpoff, 1936), and by inhalation (Syrjala et al., 1985; Teutsch et al., 1979). *F. tularensis* is highly infectious and as few as 10 bacteria can cause disease (Cross, 2000). The high infectivity and ease of dissemination of *F. tularensis* by aerosols has raised concerns about the potential use of *F. tularensis* as biological weapon (Sjostedt, 2007) and provided the rationale for the development of new tularemia therapeutics.

A major focus of *F. tularensis* research is to decipher the molecular mechanisms that contribute to *F. tularensis* pathogenesis. The strategy for identification of virulence-associated genes has largely focused on generating mutations in putative virulence genes and assessing the resultant strains for growth attenuation in a murine tularemia model. The traditional method for assessing *F. tularensis* growth and dissemination *in vivo* requires challenging large numbers of animals with the test and control *F. tularensis* strains. Thereafter, several animals are sacrificed from each group and dissected at each time point over a time course. *F. tularensis* titers are then determined in each mouse by plating serial dilutions of organ homogenates onto agar plates. This method requires large numbers of experimental animals and is laborious. In addition, the requirement for repeated animal sacrifice, dissection and tissue handling increases the potential for occupational exposure of researchers to *F. tularensis*, a category A select agent.

Technological advances in small animal imaging have made it possible to monitor in real-time the growth and dissemination of fluorescent or bioluminescent-labeled bacteria in individual animals over the entire course of infection, offering a powerful alternative to traditional methodologies. Bioluminescence has proven to be particularly useful for this application. Bioluminescence reporters have several advantages over fluorescence reporters for *in vivo* imaging studies. Luminescence reporters are more sensitive and have lower background levels, and they do not share the auto-fluorescence or signal quenching issues that limit the utility of fluorescent reporters for *in vitro* and *in vivo* imaging applications. In addition, there are

numerous methods available for the detection of bioluminescence (e.g., CCD camera, plate reader, film exposure, scintillation counter). Bioluminescence reporters have a number of applications in bacterial pathogenesis including the quantification of gene expression, virulence analysis, and the evaluation of therapeutic agents. Bioluminescence tagging vectors that express the *Photorhabdus luminescens* lux operon have been used to follow in real-time the growth and dissemination of a number of pathogens in animal models. However, to date bioluminescence reporters have not yet been published for use in *F. tularensis*.

In this work we describe the construction of a new *F. tularensis* reporter plasmid that constitutively expresses the *P. luminescens* lux operon. We show that the presence of this plasmid in type A and type B *F. tularensis* results in bioluminescence production, which could be used to follow *F. tularensis* growth and dissemination *in vitro* and *in vivo*.

C. Materials and Methods

Bacterial strains and growth conditions

F. tularensis strains LVS (live vaccine strain) and Schu S4 were obtained from the Centers for Disease Control and Prevention (CDC, Atlanta, GA). All work involving Schu S4 was performed in a CDC-approved BSL3 facility at The University of Tennessee Health Sciences Center in accordance with approved BSL3 protocols. *F. tularensis* strains were cultured in modified Mueller Hinton broth (MH broth supplemented with 10 g/L tryptone, 0.1% glucose, 0.025% ferrous pyrophosphate, 0.1% L-cysteine, and 2.5% calf serum) or on BHI-chocolate agar (BHI agar supplemented with 1% hemoglobin and 1% IsoVitalex). *Eschericia coli* strain EC100lpir (Epicentre, Madison, WI) was used as a host for the cloning experiments and was grown in Luria-Bertani (LB) broth or on LB agar at 37˚C. Antibiotics were used at the following concentrations when necessary: kanamycin (km) at 50 mg/mL for *E. coli* and 10 mg/mL for *F. tularensis*; cefprozil was used at 300 mg/mL for *F. tularensis*; carbenicillin was used at 100 mg/mL for *E. coli*.

Recombinant DNA methods

Recombinant DNA methods were performed according to standard protocols. Restriction enzymes were purchased from New England Biolabs (Beverly, MA). PCR amplification was performed using Biolase DNA polymerase (Bioline, Taunton, MA) or Pfu DNA polymerase (Stratagene, Cedar Creek, TX). *F. tularensis* was transformed by electroporation as previously described (Bina et al., 2006) except that outgrowth following electroporation was limited to one hour (for Schu S4) or two hours (LVS) before plating onto selective media.

Construction of pXB173-lux

The *Francisella-E. coli* shuttle vector pXB167 (Bina et al., 2006) was used as a starting template for construction of pXB173-lux. The initial step in construction was to replace the ColE1 origin of replication with a cassette encoding the R6K origin of replication and the conjugal origin (*oriT*) from pBSL238 (Alexeyev and Shokolenko, 1995). This was accomplished by digestion of pXB167 with AclI and PacI restriction endonucleases to remove the ColE1 origin of replication. The resulting 4,090 bp fragment was made blunt-ended by treatment with the Klenow fragment of DNA polymerase before being ligated to the 784 bp *ori*R6k and *oriT* PCR amplicon that was obtained from pBSL238 by PCR using the *oriF* (5'-CGATCTACTATGCCATGTCAGCCGTTAAGTGTTCC-3') and *oriR* (5'-GGGATATCGGGGATCAATTCCGTGATAGGTGG-3') primers to produce pXB168 (Fig. 1) to produce pXB168 (Fig.1).

We then replaced the *gfp* gene in pXB168 with the *aph3'* gene that encoded kanamycin resistance. This was accomplished by digestion of pXB168 with BamHI and ClaI restriction

enzymes to remove the *gfp* allele. The resulting 4,119 bp fragment was rendered blunt-ended by treatment with Klenow fragment before being ligated to the 901 bp kanamycin resistance gene (*aph3'*) was obtained from TN:EZ by PCR using the (Epicentre, Madison, WI) using the KanF (5'-AAGGCGCGCCACGCGTAGGAGTTTGTTATGAGCCATATTCAACGGGAA-3') and KanR (5'-GCACGCGTCAAGTCAGCGTAATGCTCTGCCAG-3') primers to generate pXB169. pJB173-lux was then generated by digestion of pXB169 with KpnI and HinCII restriction enzymes to remove the *shv-2* allele. The resulting 3,987 bp fragment was then ligated with the 5.9 kb lux operon that was derived from digestion of pXB128-lux (Bina laboratory collection) with KpnI and XmaI restriction enzymes. The results of this ligation placed the lux operon downstream and in the same orientation as the constitutively expressed Francisella *gro* promoter (indicated as P_{gro} in **Fig. 1**). The DNA sequence of pXB173-lux was confirmed by DNA sequencing at the Molecular Resource Center of The University of Tennessee Health Science Center (Memphis, TN). The annotated DNA sequence is provided in the supplementary data. pXB173-lux has been assigned the Genbank accession number HM017829.

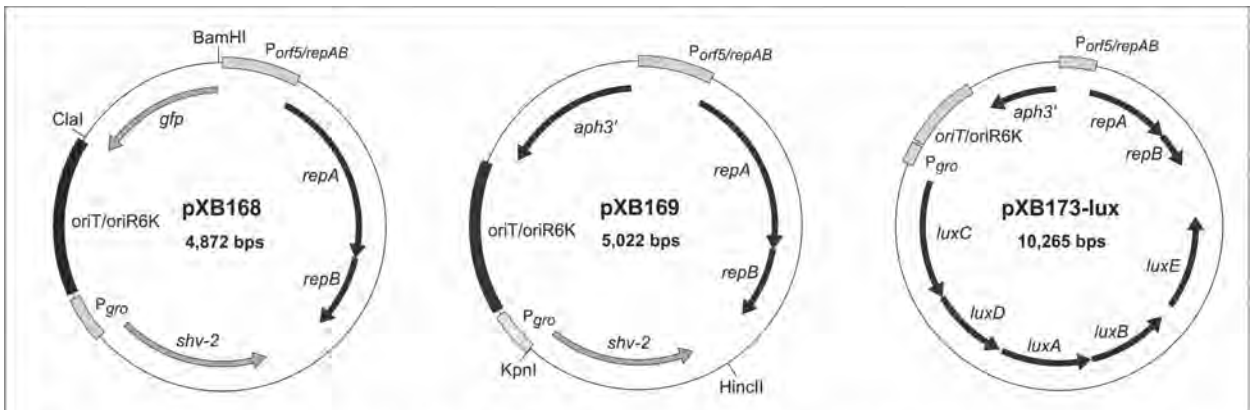


Fig.1. Construction of pXB173-lux. Plasmid pXB168 is derived from pXB167 (Bina et al., 2006). Only relevant restriction sites are shown. The details for construction are given in the text.

Bioluminescence detection and animal challenge studies

The limits of bioluminescence detection of LVS-pXB173-lux was determined by making serial LVS-pXB173 culture dilutions in white 96-well microtiter plates. Bioluminescence production was then quantified by use of an IVIS Spectrum imaging system (Caliper Life Sciences, Hopkinton, MA) according to the manufacturer's directions. The utility of pXB173-lux in *F. tularensis* was assessed in a murine intranasal challenge model as previously described (Lavine et al., 2007). Briefly, 12 week-old female BALB/c mice were challenged intranasally with ~$5x10^5$ CFU of *F. tularensis* LVS-pXB173-lux in a total volume of 50 µl that was administered as 25 µl per naris. Bioluminescence was then used as a reporter to following bacterial dissemination starting at 3 hrs post challenge and then at 24-hour intervals until the conclusion of the experiment. Bioluminescence production in the mice was quantified by use of the IVIS Spectrum imaging system according to the manufacturer's directions.

Bioluminescence production in *F. tularensis* Schu S4 was assessed by use of the IVIS Spectrum to image an agar plate that had been inoculated with both the LVS and Schu S4

strains of *F. tularensis* bearing the pXB173-lux vector. Due to BSL3 restrictions we are currently unable to image mice infected with *F. tularensis* Schu S4 on the IVIS imaging system.

***Plasmid stability* determination**

The stability of pXB173-lux in *F. tularensis* Schu S4 was assessed *in vitro* as follows. A fresh overnight culture of *F. tularensis* Schu S4-pXB173 was successively cultured in MH broth without Km for four days. The presence of the plasmid was then determined by plating serial dilutions of the culture at days 1 and 4 onto BHI-chocolate agar plates with and without Km. The ratio of Km-resistant to Km-sensitive bacteria was then calculated to determine pXB173-lux stability in the absence of antibiotic selection. In vivo stability of pXB173-lux was determined by intranasal infection of mice with ~10^3 cfu of *F. tularensis* Schu S4-pXB173. The spleen was then collected from one mouse on days 1, 5 and 6 and homogenized in 1 ml of PBS before 0.1 ml of disruption buffer (2.5% saponin, 15% BSA, in PBS) was added with light vortexing. Serial dilutions of the spleen homogenates were then plated onto BHI-chocolate agar plates with and without Km. The ratio of Km-resistant to Km-sensitive bacteria was then calculated as an indicator of *in vivo* plasmid stability.

D. Results and Discussion

Construction of pXB169

We previously described the construction of three shuttle vectors (pXB136, pXB160 and pXB167) (Bina et al., 2006) that were derived from pFNLTP6gfp (Maier et al., 2004). In electroporation experiments we observed that pXB136 could be efficiently transformed into Schu S4 with selection for cefprozil resistance, however, we were not able to recover Schu S4 transformants when selecting for kanamycin resistance (data not shown). Since the kanamycin resistance locus in pXB136 is derived from the pFNLTP shuttle vectors, this observation is consistent with the previous finding that pFNLTP-based vectors transformed poorly into Schu S4 (LoVullo et al., 2006) and suggested that the poor transformation and plasmid instability observed in pXB136 was likely due to inefficient expression of the kanamycin resistance allele in type A *F. tularensis* strains. In silico analysis of pXB136 suggested that the *repAB* locus in pXB136 (and in the pFNLTP vectors) contained a divergently transcribed promoter, denoted as P_{orf5} in Fig. 1, located 808 bp upstream of the kanamycin resistance gene (i.e. *aph3*'). Downstream of the P_{orf5} is *orf5'* which encodes a truncated gene that was hypothesized to form part of a two component toxin-antitoxin system that was present in the parent plasmid pFNL10 (Pavlov et al., 1996). Downstream of *orf5'* was the f1 origin of replication and the *aph3'* gene which originated from pCR2.1-TOPO (Maier et al., 2004). As P_{orf5} was derived from pFNL10, we hypothesized that it likely encoded an active *F. tularensis* promoter and contributed to *aph3'* expression in pXB136 and that the intervening 808 bp sequence inhibited *aph3'* expression in Schu S4. To test this hypothesis we deleted the 808 bp intervening region. The resulting plasmid retained kanamycin resistance in *E. coli* and LVS and gained the ability to be retained by Schu S4. The resulting plasmid transformed into Schu S4 at an efficiency that was equivalent to what was observed with LVS (~10^5 cfu/mg DNA). Collectively these results suggested that the *repAB* promoter region contained a divergently transcribed promoter that was constitutively expressed in both *E. coli* and *F. tularensis*. It is unclear why pXB136 and the pFNLTP plasmids display different stabilities in LVS and Schu S4 with selection for kanamycin resistance.

Construction of pXB173-lux

Having established that the Schu S4 stability problems associated with our previous vectors was likely due to expression of the kanamycin resistance allele and not some inherent problem with the plasmid construct, we set out to design a new shuttle vector that could be used as a bioluminescence reporter in *F. tularensis*. To construct the bioluminescence reporter plasmid we first replaced the high copy number origin of replication that was present in pXB167 with a low copy number origin of replication and a conjugal origin of transfer. This was accomplished by replacement of the pXB167 ColE1 origin of replication with a cassette that encoded the R6K origin of replication and the RP4 origin of transfer to generate pXB168. This effectively reduced the plasmid copy number in *E. coli* and introduced an origin of transfer to facilitate conjugal transfer of the plasmid into *F. tularensis*. Conjugation represents a very efficient and easy method for introduction of plasmids into *F. tularensis*. The *shv-2* marker was then replaced with the *aph3'* allele as kanamycin resistance is the most reliable and widespread genetic marker used in type A *F. tularensis* strains (e.g. Schu S4). We used the *orf5* promoter to drive expression of *aph3'* so that we could use the P_{gro} promoter to drive expression of the lux reporter construct (see below). The resulting plasmid, pXB169, was transformed into Schu S4 with high efficiency (~10^5 transformants per mg/DNA).

The bioluminescence reporter plasmid pXB173-lux (**Fig. 1**) was then generated from pXB169 by cloning the *P. luminescens* lux operon downstream of the *F. tularensis gro* promoter. The *P. luminescens* lux operon contains the genes that are required for production of both luciferase (*luxAB*) and luciferin (*luxCDE*) and expression of the lux operon results in concomitant light production. Since the *F. tularensis gro* promoter is constitutively expressed in *E. coli* and *F. tularensis*, the presence of pXB173-lux in *E. coli*, LVS and Schu S4 results in constitutive bioluminescence production as observed in **Fig. 2**. The *in vitro* detection limit for LVS-pXB173-lux in white 96-well microtiter plates

Fig. 2. Bioluminescence production by *F. tularensis*. Overnight cultures of *F. tularensis* containing pXB173-lux (LVS on the upper half of plate; Schu S4 on the lower half of plate) were inoculated onto the surface of a modified Mueller Hinton agar plate using a Dacron-tipped swab and incubated at 37C for 18 hrs when the plate was imaged for bioluminescence production using an IVIS Spectrum imaging system. Photon emission intensity is represented as a pseudocolor image that is superimposed onto the surface of the inoculated agar plate.

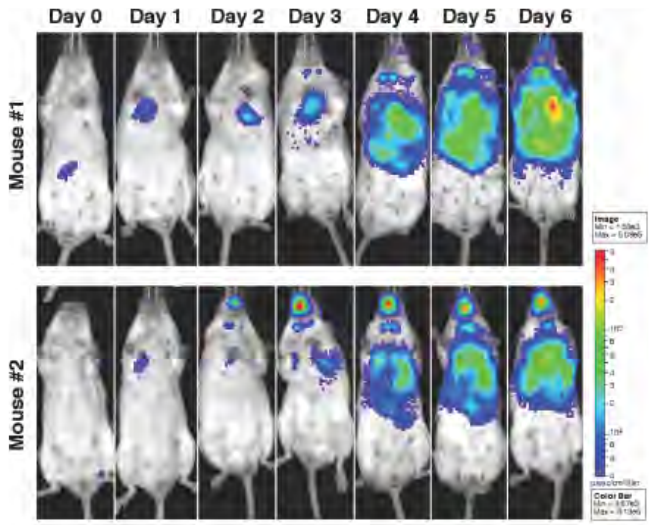


Fig. 3. Visualization of *F. tularensis* LVS-pXB173-lux in mice by bioluminescence imaging. Two twelve week-old BALB/c mice were challenged with 5 x 10^5 CFU *F. tularensis* LVS-pXB173-lux in a total volume of 50 µl of PBS via the intranasal route. Bioluminescence production in the mice was then visualized using an IVIS Spectrum Imaging system at 24-hour intervals beginning 3 hours after administration of the challenge dose. Exposure times varied based on bioluminescent signal intensities in an effort to collect between 600 and 60,000 counts, and image scaling was normalized by converting total counts to photons/second. Results shown here are representative of several experiments of similar design.

was ~2,000 cfu, which suggests that pXB173-lux likely can be used to follow *F. tularensis* growth in cell culture studies.

Use of pXB173-lux to follow F. tularensis dissemination in mice.
We documented the utility of pXB173-lux by testing whether it could be used as a reporter to follow *F. tulanensis* growth in a murine model of tularemia in real-time. We therefore challenged two BALB/c mice with ~10^5 cfu of LVS-pXB173-lux by the intranasal route (**Fig. 3**). In the first mouse, the majority of the LVS inoculum was detected in the stomach at three hours post challenge, suggesting that at least a portion of the intranasal challenge dose failed to reach the lungs and was swallowed. Twenty four-hours later, the bioluminescence production in the stomach of mouse 1 had resolved and LVS was clearly visualized in the lungs of both animals and in the upper respiratory tract of mouse two. The upper airway infection intensified over the six day course of the experiment in mouse two. It is unclear whether this upper respiratory tract infection occurs in natural inhalation infections or is an artifact of the intranasal inoculation method that is widely used by the tularemia research community. The images also showed that LVS disseminated to the cervical lymph nodes of mouse one (day 3) and mouse two (day 2) and that colonization of the lymph nodes intensified throughout the study period. Beginning on day 3 post-challenge LVS was observed in the liver of both mice, and by day 4 post-challenge, the livers of both mice were heavily colonized. On day 6 the both mice exhibited extensive bacterial dissemination, which correlated with other signs of severe tularemic disease (i.e. significant weight loss, ruffled fur and reduced physical activity) and the experiment was terminated. These results validate that pXB173-lux can be used as a reporter to follow *F. tularensis* dissemination in mice.

The stability of pXB173-lux in *F. tularensis* Schu S4 was assessed to validate the use of this plasmid in a Type A strain background. The plasmid was well maintained *in vitro* with 82% of the bacteria retaining the plasmid following growth for four successive subcultures in the absence of antibiotic selection. *In vivo* stability was similar to the *in vitro* results with 84% and 76% and of the bacteria retaining the plasmid on days 5 and 6, respectively. This demonstrates that pXB173-lux is stable in *F. tularensis* Schu S4, while the data presented in **figure 2** demonstrate that the lux reporter works in *F. tularensis* Schu S4. Although we are currently restricted from performing imaging experiments with *F. tularensis* Schu S4-infected mice, our collective results strongly suggest that our results with LVS-pXB173-lux will extend to use of this reporter for in vivo studies with Type A *F. tularensis* strains.

E. Key Research Accomplishments

a. Generated a luminescence reporter for *F. tularensis*
b. Validated use of the reporter *in vivo* and *in vitro* in both Type A and Type B *F. tularensis*
c. Determined limits of detection in microtiter plates and in mice

F. Reportable Outcomes:

Publication

1. Bina XR, Miller MA, Bina JE. Construction of a bioluminescence reporter plasmid for *Francisella tularensis*. Plasmid. 2010 Nov;64(3):156-61., PMID: 20620161, PMCID: PMC2943566 [Available on 2011/11/1] (see appendices for pdf)

G. Conclusion

The results presented above show that bioluminescence is a highly sensitive reporter that can be used to follow *F. tularensis* growth in mice in real time. Bioluminescence represents a new tool for the tularemia research community that has not been previously available. In particular, the use of pXB173-lux can greatly facilitate animal and cell culture studies with virulent type A *F. tularensis* strains. As most analysis previously depended on terminal end point assays, the use of bioluminescence should greatly reduce both the labor cost and number of animals that are required for these assays while limiting the potential for occupational exposure of researchers to a potentially fatal pathogen.

H. References

1 Alexeyev, M. F., Shokolenko, I. N., 1995. RP4 oriT and RP4 oriT-R6K oriV DNA cassettes for construction of specialized vectors. Biotechniques. 19**,** 22-4, 26.

2 Anda, P., et al., 2001. Waterborne outbreak of tularemia associated with crayfish fishing. Emerging Infectious Diseases. 7**,** 575-82.

3 Bina, X. R., et al., 2006. The Bla2 beta-lactamase from the live-vaccine strain of *Francisella tularensis* encodes a functional protein that is only active against penicillin-class beta-lactam antibiotics. Arch Microbiol. 186**,** 219-28.

4 Cross, T. J. a. R. L. P., 2000. *Francisella tularensis* (tularemia). In: G. L. Mandell, Bennett J.E. and R. Dolin, (Ed.), Principles and Practice of Infectious Diseases. Churchill Livingston, Philadelphia.

5 Evans, M. E., 1985. *Francisella tularensis*. Infection Control. 6**,** 381-3.

6 Francis, E., 1937. Sources of infection and seasonal incidence of tularemia in man. Public Health Reports. 52**,** 103.

7 Greco, D., et al., 1987. A waterborne tularemia outbreak. European Journal of Epidemiology. 3**,** 35-8.

8 Karpoff, S. P., and N.I. Antononoff., 1936. The spread of tularemia through water as a new factor in its epidemiology. Journal of Bacteriology. 32**,** 243.

9 Lavine, C. L., et al., 2007. Immunization with heat-killed *Francisella tularensis* LVS elicits protective antibody-mediated immunity. Eur J Immunol. 37**,** 3007-20.

10 LoVullo, E. D., et al., 2006. Genetic tools for highly pathogenic *Francisella tularensis* subsp. tularensis. Microbiology. 152**,** 3425-35.

11 Maier, T. M., et al., 2004. Construction and characterization of a highly efficient Francisella shuttle plasmid. Appl Environ Microbiol. 70**,** 7511-9.

12 Pavlov, V. M., et al., 1996. Cryptic plasmid pFNL10 from *Francisella novicida*-like F6168: the base of plasmid vectors for Francisella tularensis. FEMS Immunol Med Microbiol. 13**,** 253-56.

13 Sjostedt, A., 2007. Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. Ann N Y Acad Sci. 1105**,** 1-29.

14 Syrjala, H., et al., 1985. Airborne transmission of tularemia in farmers. Scandinavian Journal of Infectious Diseases. 17**,** 371-5.

15 Tarnvik, A., 1989. Nature of protective immunity to *Francisella tularensis*. Reviews of Infectious Diseases. 11**,** 440-51.

16 Teutsch, S. M., et al., 1979. Pneumonic tularemia on Martha's Vineyard. New England Journal of Medicine. 301**,** 826-8.

1.1.2 In vivo studies

A. Overview

We have been unable to identify a challenge dose of type A *Francisella tularensis* (FT Schu S4) that yields a significantly different outcome to infection in B6 and D2 mice. Both strains succumb to infection at essentially the same rate no matter what route of infection is used. Therefore, we will need to identify other readouts for phenotypic differences between the two parental strains that can be used for QTL analysis in BXD mice. We have completed a timecourse study in which B6 and D2 mice were challenged intranasally with approximately 250 CFU of WT FT Schu S4 and then mice were sacrificed (5/group/timepoint) at 12, 18, 24, and 48 hrs post-infection. Based on studies using FT LVS, we expected to see a difference in cell recruitment to the lungs 18-24 hours post-infection. However, we observed no significant differences in cell recruitment to the lungs at any of the timepoints (data not shown). We posit that in order to observe a difference in cell recruitment, we will have to challenge mice with a much larger dose of FT (the studies with LVS were performed with 5 x 10^4 CFU). As a component of the timecourse study that we recently completed, we collected BALF and serum that will be subjected to multiplex cytokine measurements using a Luminex 32-plex cytokine/chemokine kit (these studies are scheduled for next week.

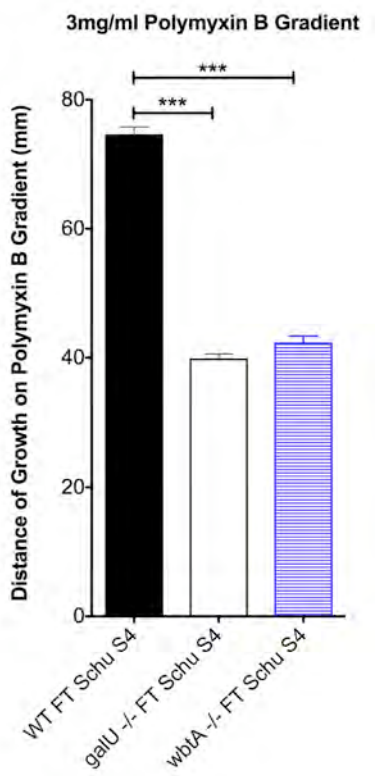


Figure 1. *galU* and *wbtA* mutant strains of FT Schu S4 exhibit increased sensitivity to polymyxin B. WT, Δ*galU*, and Δ*wbtA* strains of FT Schu S4 were swabbed from an overnight culture onto Mueller-Hinton agar plates containing a continuous 0-3 mg/ml polymyxin B gradient using a soaked Q-tip. Sensitivity of each FT strain to polymyxin B was determined by measuring the distance of growth of each strain on the gradient. Statistical analysis was performed via One-Way ANOVA using a Bonferonni's multiple comparison post-test (p value <0.0001).

We have continued our efforts to identify essential virulence determinants of FT Schu S4. In our progress report from September 2010, we described work we had performed using a *galU* mutant strain of FT LVS as well as some work with a *wbtA* transposon mutant strain made in the FT Schu S4 background showing that these two genes appear to be critical for full virulence of FT LVS and FT Schu S4, respectively. We have recently recovered clean deletion mutants of both *galU* and *wbtA* in the Schu S4 background. As expected, both of these mutant strains display heightened sensitivity to polymyxin B (**Figure 1**), confirming that the outer envelope of these two mutant strains have been altered. To determine whether one or both of these mutants are attenuated for virulence *in vivo*, C57Bl/6J mice (4-5/group) were challenged intranasally with titrated numbers (5, 50, 500, and 5,000 CFU) of the *galU* or the *wbtA* mutant strains of FT Schu S4 and then mice were monitored for weight retention and survival for 20 days. While each of the challenged mice became infected (as evidenced by their transient weight loss, **Figure 2A**), the wbtA mutant strain displayed a significantly attenuated virulence phenotype as evidenced by the survival of each mouse that received a 500 CFU or smaller dose, and and survival of 2/4 mice that received a challenge with 5000 CFU (**Figure 2B**). By comparison, all mice that received 50CFU of WT FT Schu S4 succumbed to infection within 5 days of challenge. Surprisingly, the galU mutant strain displayed wild-type virulence characteristics (data not shown).

To determine whether the attenuated wbtA mutant strain of FT has potential utility as a vaccine strain, the mice that survived challenge with the wbtA mutant strain were challenged on day 70 with 50 CFU of WT FT Schu S4. Unfortunately, this mutant strain failed to elicit a protective anamnestic response, leaving the mice unprotected against virulent FT challenge (**Figure 2C**).

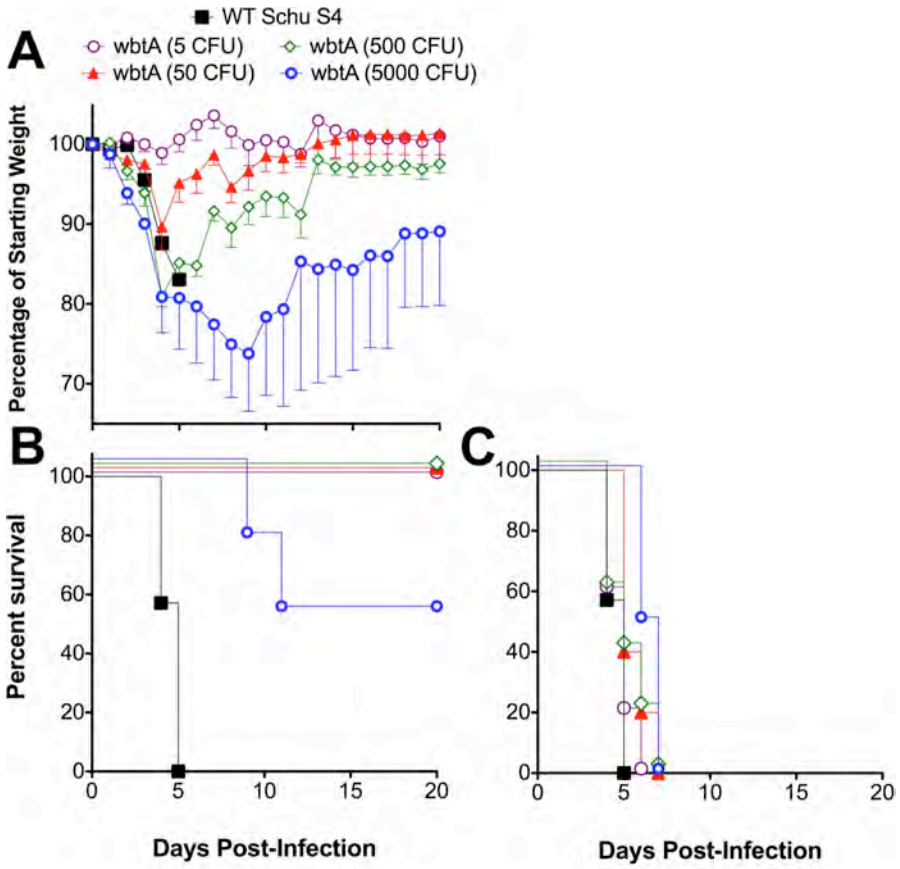


Figure 2. A wbtA clean deletion mutant is significantly attenuated for virulence, but it fails to elicit protective immunity. C57Bl.6J mice were challenged with the indicated doses of either FTΔwbtA or with 50 CFU WT FT and then monitored for weight retention (Panel A) and survival (**Panel B**) for 20 days. 50 days later, the mice that survived challenge with the wbtA mutant strain were challenged with 50 CFU of WT FT and monitored for survival (**Panel C**).

We have also continued our analysis of a series of other clean deletion mutants of FT Schu S4 that were mentioned in the previous progress report (ΔFTT1188, ΔFTT 0584, and ΔFTT0453). Each of these genes are analogous to genes that were found to be required for virulence of *F. novicida* [14], and as such were considered strong candidates as virulence determinants of FT Schu S4. This hypothesis has turned out to be incorrect. FT Schu S4ΔFTT1188 and FT Schu S4ΔFTT0584 were analyzed for virulence in mice and were found to be essentially as virulent as WT FT Schu S4 (data not shown). In addition, we found that both of these mutant strains had WT replication characteristics in both RAW264.4 and THP-1 cells (data not shown). Likewise, the FT Schu S4ΔFTT0453 and ΔFTT0941 strains retained WT virulence characteristics (data not shown). The FTT0941 mutant strain displayed no growth phenotype in either RAW264.4 or THP-1 cells (data not shown). Interestingly, the FTT0453 mutant strain displayed WT growth characteristics in RAW264.4 cells, but has a significant growth defect in the semi-activated THP-1 cells. This may indicate that the FTT0453 mutant is not well adapted for growth within, or is killed easier by, activated macrophages. Despite this growth defect *in vitro*, the FTT0453 mutant strain retained full virulence.

B. Key Research Accomplishments

- **a.** Initiated studies to identify phenotypic differences between B6 and D2 mice following challenge with virulent FT Schu S4.
- **b.** Created a highly attenuated mutant strain (∆wbtA) of FT Schu S4.
- **c.** Created and produced a series of mutation strains of FT Schu S4. Each of the mutations has been shown to be highly attenuating in either (or both) the LVS and FT *novicida* backgrounds. We have shown that of the mutant strains tested, only the wbtA mutant strain displays an attenuated phenotype for disease in vivo. These findings have galvanized our belief that the LVS and FT novicida infection models do not correlate well with the highly virulent type A FT murine model.

C. Reportable Outcomes

Publications:

1. Bina XR, Miller MA, Bina JE. **2010**. Construction of a bioluminescence reporter plasmid for *Francisella tularensis*. Plasmid. Nov;64(3):156-61., PMID: 20620161, PMCID: PMC2943566 [Available on 2011/11/1] (see appendices for pdf)
2. Himangi R. Jayakar, Jyothi Parvathareddy, Elizabeth A. Fitzpatrick, Xiaowen R. Bina[1], James E. Bina, Fabio Re, and Mark A. Miller. **2011**. A *galU* Mutant of *Francisella tularensis* is Attenuated for Virulence in a Murine Pulmonary Model of Tularemia. BMC Microbio (accepted with revisions).

Manuscript in preparation

1. Mark A. Miller, Jennifer M. Stabenow, Jyothi Parvathareddy, Andrew J. Wodowski, Thomas P. Fabrizio, Xiaowen R. Bina**,** Lillian Zalduondo, and James E. Bina. **2011**. Visualization of Murine Intranasal Dosing Efficiency Using Luminescent *Francisella tularensis*: Effect of Instillation Volume and Form of Anaesthesia. J Microbio Methods.

D. Conclusions

We have initiated studies to identify phenotypic differences between FT disease-state B6 and D2 mice. We expect to have significant progress with this line of investigation over the next few months. We have continued to evaluate mutant strains of FT Schu S4 in an effort to identify critical virulence factors and to work towards creation of a well-characterized attenuated strain that could be used to elicit protective responses to type A stains of *Francisella tularensis*. Consistently we are finding that genes that have been shown to be required for full virulence in the *F. novicida* background are not required for full virulence of the type A Schu S4 strain. These findings have underscored the importance of performing these types of studies with the highly virulent type A strains.

We have recently created two high-priority clean in-frame deletion mutant of FT Schu S4 (∆*galU* and ∆*wbtA*). One of these mutants (∆wbtA) was highly attenuated, while the other retained WT virulence *in vivo*. Unfortunately, the ∆wbtA strain of FT fails to elicit long-lived protective immunity to WT FT challenge and, therefore, has little potential to serve as a vaccine strain.

E. References

1. N Harris, LL Peters, EM Eicher, M Rits, D Raspberry, QG Eichbaum, M Super, RA Ezekowitz: **The exon-intron structure and chromosomal localization of the mouse macrophage mannose receptor gene Mrc1: identification of a Ricin-like domain at the N-terminus of the receptor**. *Biochem Biophys Res Commun* 1994, **198**:682-92.
2. T Gruber, F Fresser, M Jenny, F Uberall, M Leitges, G Baier: **PKCtheta cooperates with atypical PKCzeta and PKCiota in NF-kappaB transactivation of T lymphocytes**. *Mol Immunol* 2008, **45**:117-26.
3. N Hermann-Kleiter, N Thuille, C Pfeifhofer, T Gruber, M Schafer, C Zitt, A Hatzelmann, C Schudt, M Leitges, G Baier: **PKCtheta and PKA are antagonistic partners in the NF-AT transactivation pathway of primary mouse CD3+ T lymphocytes**. *Blood* 2006, **107**:4841-8.
4. A Hartner, I Marek, N Cordasic, C Haas, H Schocklmann, G Hulsmann-Volkert, I Plasa, W Rascher, KF Hilgers, K Amann: **Glomerular regeneration is delayed in nephritic alpha 8-integrin-deficient mice: contribution of alpha 8-integrin to the regulation of mesangial cell apoptosis**. *Am J Nephrol* 2008, **28**:168-78.
5. B Bieritz, P Spessotto, A Colombatti, A Jahn, F Prols, A Hartner: **Role of alpha8 integrin in mesangial cell adhesion, migration, and proliferation**. *Kidney Int* 2003, **64**:119-27.
6. J Martin, F Magnino, K Schmidt, AC Piguet, JS Lee, D Semela, MV St-Pierre, A Ziemiecki, D Cassio, C Brenner, et al: **Hint2, a mitochondrial apoptotic sensitizer down-regulated in hepatocellular carcinoma**. *Gastroenterology* 2006, **130**:2179-88.
7. HB Eberle, RL Serrano, J Fullekrug, A Schlosser, WD Lehmann, F Lottspeich, D Kaloyanova, FT Wieland, JB Helms: **Identification and characterization of a novel human plant pathogenesis-related protein that localizes to lipid-enriched microdomains in the Golgi complex**. *J Cell Sci* 2002, **115**:827-38.
8. A Schebesta, S McManus, G Salvagiotto, A Delogu, GA Busslinger, M Busslinger: **Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function**. *Immunity* 2007, **27**:49-63.
9. GR Moore, CS Raine: **Immunogold localization and analysis of IgG during immune-mediated demyelination**. *Lab Invest* 1988, **59**:641-8.
10. S Aguera-Gonzalez, P Boutet, HT Reyburn, M Vales-Gomez: **Brief residence at the plasma membrane of the MHC class I-related chain B is due to clathrin-mediated cholesterol-dependent endocytosis and shedding**. *J Immunol* 2009, **182**:4800-8.
11. DA Alvarez Arias, N McCarty, L Lu, RA Maldonado, ML Shinohara, H Cantor: **Unexpected role of clathrin adaptor AP-1 in MHC-dependent positive selection of T cells**. *Proc Natl Acad Sci U S A*, **107**:2556-61.
12. Y Takahama, K Ohishi, Y Tokoro, T Sugawara, Y Yoshimura, M Okabe, T Kinoshita, J Takeda: **Functional competence of T cells in the absence of glycosylphosphatidylinositol-anchored proteins caused by T cell-specific disruption of the Pig-a gene**. *Eur J Immunol* 1998, **28**:2159-66.
13. M Bessler, V Rosti, Y Peng, G Cattoretti, R Notaro, S Ohsako, KB Elkon, L Luzzatto: **Glycosylphosphatidylinositol-linked proteins are required for maintenance of a normal peripheral lymphoid compartment but not for lymphocyte development**. *Eur J Immunol* 2002, **32**:2607-16.
14. DS Weiss, A Brotcke, T Henry, JJ Margolis, K Chan, DM Monack: **In vivo negative selection screen identifies genes required for Francisella virulence**. *Proc Natl Acad Sci U S A* 2007, **104**:6037-42.

1.2 Studies with *Burkholderia pseudomallei*

1.2.1 In vivo Studies

A. Overview

On September 22, 2010, we received all 18 of the *Burkholderia pseudomallei* (Bp) strains that we ordered. We learned at about this time that Dr. Colin Manoil (U. of Washington) was planning to make a comprehensive mutant library of Bp strain 1026b. Because of the impending availability of this important resource, we believe that the 1026b strain will become the most well characterized reference strain of Bp. Therefore we decided to use this strain for BXD studies.

Our initial goal was to identify one or more disease-state phenotypes of differential susceptibility to Bp1026b in the two parental strains (B6 and D2). To determine if there was a difference in susceptibility between the parental strains, we challenged B6 and D2 mice with either 50 or 100 CFU of Bp1026b via the intranasal route. We then evaluated the mice for illness via daily observation and body weight monitoring for 20 days. A clear phenotypic difference was observed between the two parental strains. All of the mice lost weight initially, confirming that they had all been infected with Bp (**Figure 1**). A clear phenotypic difference between the parentals was observed as each of the B6 mice survived the challenge while all of the D2 mice succumbed to infection by day 4 post-challenge.

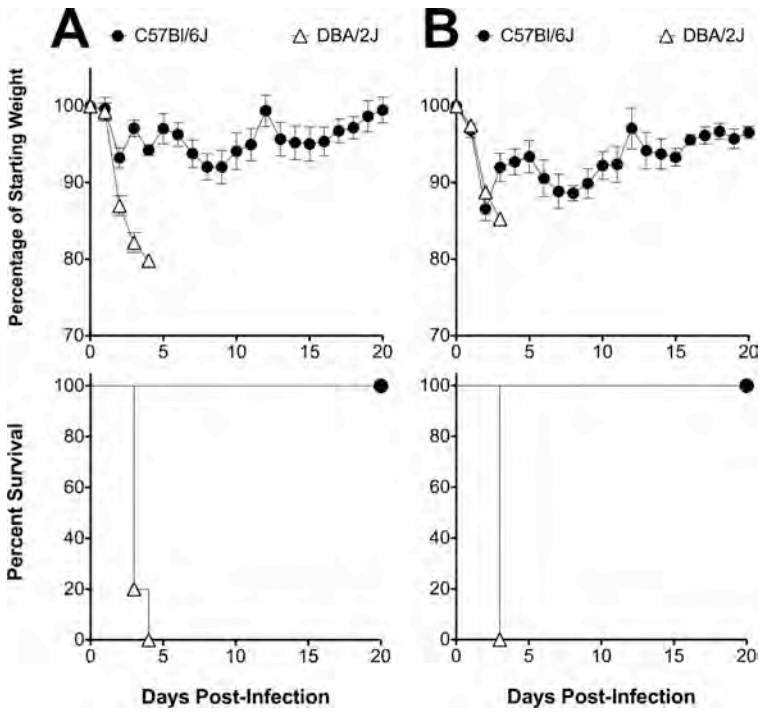


Figure 1. Weight loss and survival of C57Bl/6J and DBA/2J mice following challenge with *Burkholderia pseudomallei* strain 1026b. Mice (5/group) were challenged with either 50 (Panel A) or 100 (Panel B) CFU Bp1026b via the intranasal route. The mice were monitored daily for 20 days post-infection. Results shown here are representative of three experiments of similar design.

Based on these findings, we have initiated studies in BXD mouse strains using survival and weight loss as phenotypes of differential susceptibility to Bp1026b-mediated disease. We have now screened 33 BXD strains using group sizes of between 3 and 11 mice/strain. We are challenging the mice with 50-100 CFU of Bp1026b via the intranasal route and scoring the mice for survival and weight loss. We have analyzed the results to date using WebQTL software (www.genenetwork.org). These analyses have identified two suggestive QTLs, one that encoded on chromosome 4 and another that is encoded on chromosome 7 (**Figure 2A**). Because several of the BXD strains tested displayed an intermediate phenotype, (**Figures 2B and 2C**), it is not surprising that our preliminary results suggest that the survival phenotype is a complex trait that requires covariance of multiple host genes.

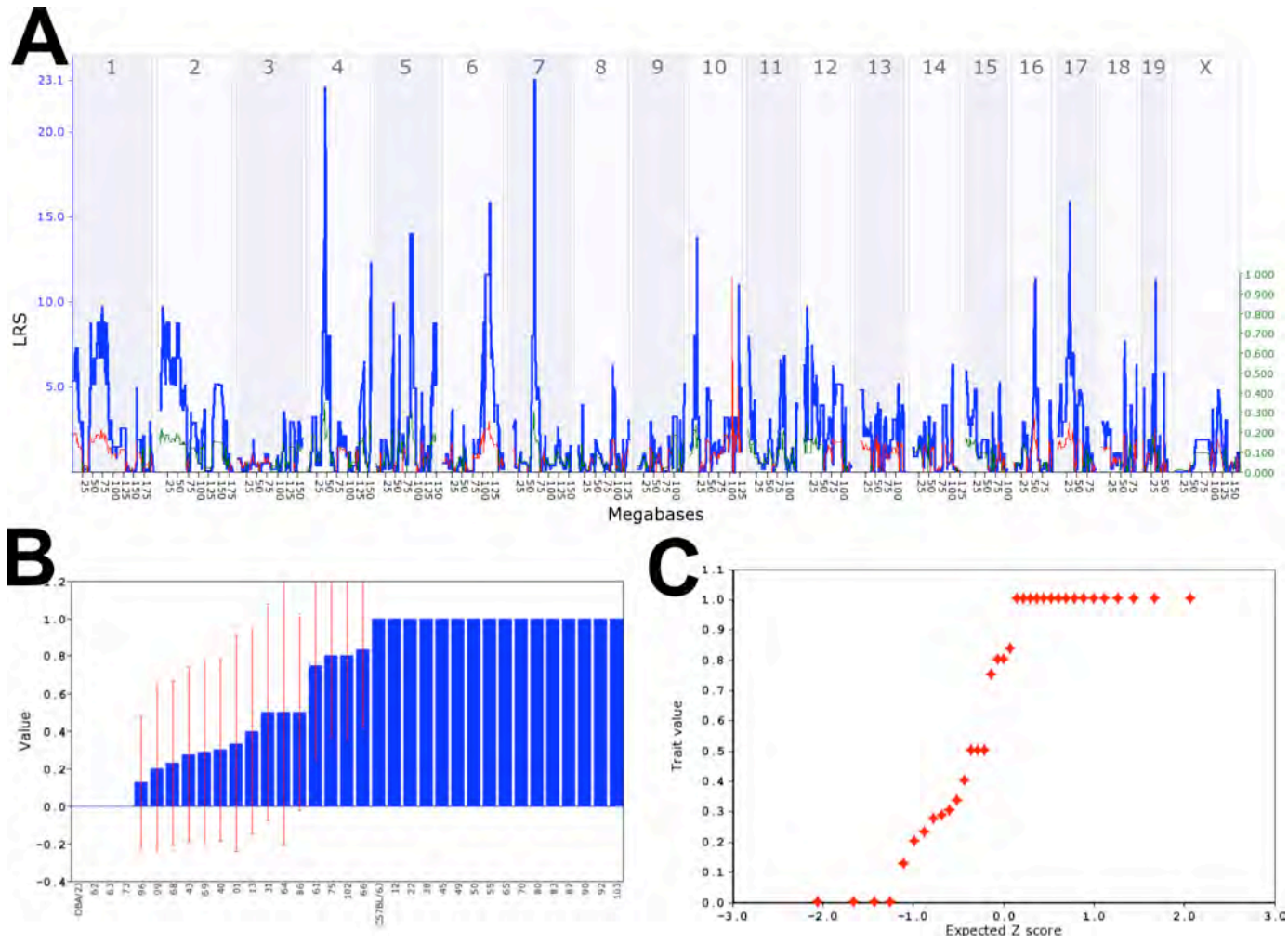


Figure 2. Preliminary QTL analysis of BXD survival phenotypes following challenge with *Burkholderia pseudomallei* strain 1026b. Interval mapping of survival phenotypes observed in 33 BXD strains (**Panel A**) was performed and is indicated by a blue line representing the LRS score across the entire mouse genome. Two tall peaks that are considered to be highly suggestive QTL were observed in chromosomes 4 and 7. Case ranking (**Panel B**) and a probability plot (**Panel C**) show the survival phenotype of each BXD strain studied to. All data analysis was performed using GeneNetwork WebQTL software.

Because the DBA/2J (as well as most susceptible BXD mice) succumb to infection by day 4 post-challenge, we predict that the differential susceptibility of the parental strains involves genes that encode components of the innate immune response or housekeeping genes that are involved in maintaining integrity of host tissues during infection. We have performed a cursory evaluation of host loci that fall within the suggestive QTLs on chromosomes 4 (**Figure 3**) and 7 (**Figure 4**). Several of the genes located within the chromosome 4 QTL interval have a known role in immune responsiveness. For instance, *pax5* encodes a transcription factor that has an important role in B-cell development. It represses B lineage-inappropriate genes and activates B cell-specific genes in B lymphocytes [8]. The *clta* gene product has been implicated in receptor-mediated phagocytosis of debris via clathrin-coated pits on the surface of macrophages in certain disease process [9]. It has also been shown to play a role in subcellular localization and trafficking of MHC class I-related chain molecules on the surface of target cells, which subsequently leads to lysis by immune effector cells [10, 11]. *Pigo* encodes a component that is essential for the initial steps of glycosylphosphatidylinositol (GPI)-anchor biosynthesis [12]. As such, GPI-linked proteins have been shown to be a requirement for normal lymphocyte function

and for maintaining normal peripheral lymphoid homeostasis [13]. *The hint2* gene product defines a novel class of mitochondrial apoptotic sensitizers [6] that have been shown to influence mitochondria-dependent apoptosis. *Glipr2* (also known as GAPR-1) is highly expressed in the lungs and peripheral leukocytes [7].

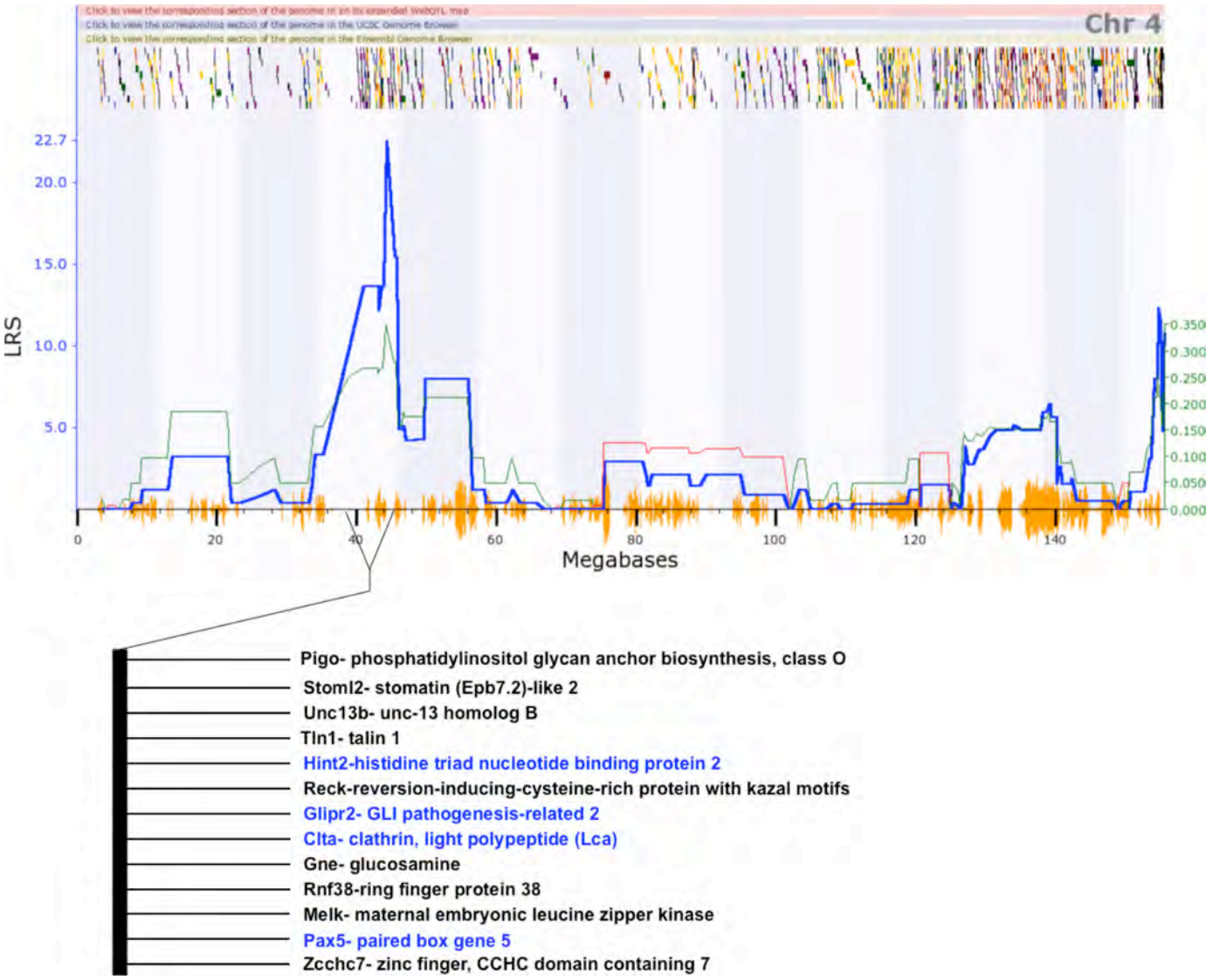


Figure 3. Identity of host genes encoded within the suggestive QTL on chromosome 4. The interval map for chromosome 2 is shown in the top panel. A list of genes encoded within the highly suggestive QTL is shown in the bottom panel. Genes that are potentially involved in immune processes are listed in blue.

We have also examined weight loss at day 1, day 2, and day 3 post-infection as traits for QTL analysis. WebQTL analysis of the day-3 time point reveals one suggestive QTL in chromosome 12 and three additional peaks (in chromosome 1, chromosome 2, and chromosome 8) that are nearing suggestive QTL status. The peak observed in chromosome 2 appears to overlap the highly suggestive QTL that was identified using survival as a readout (data not shown). We are

hopeful that the analysis of additional BXD strains will lead to statistically relevant QTL mapping using weight loss at day 3 post-infection as a phenotypic readout.

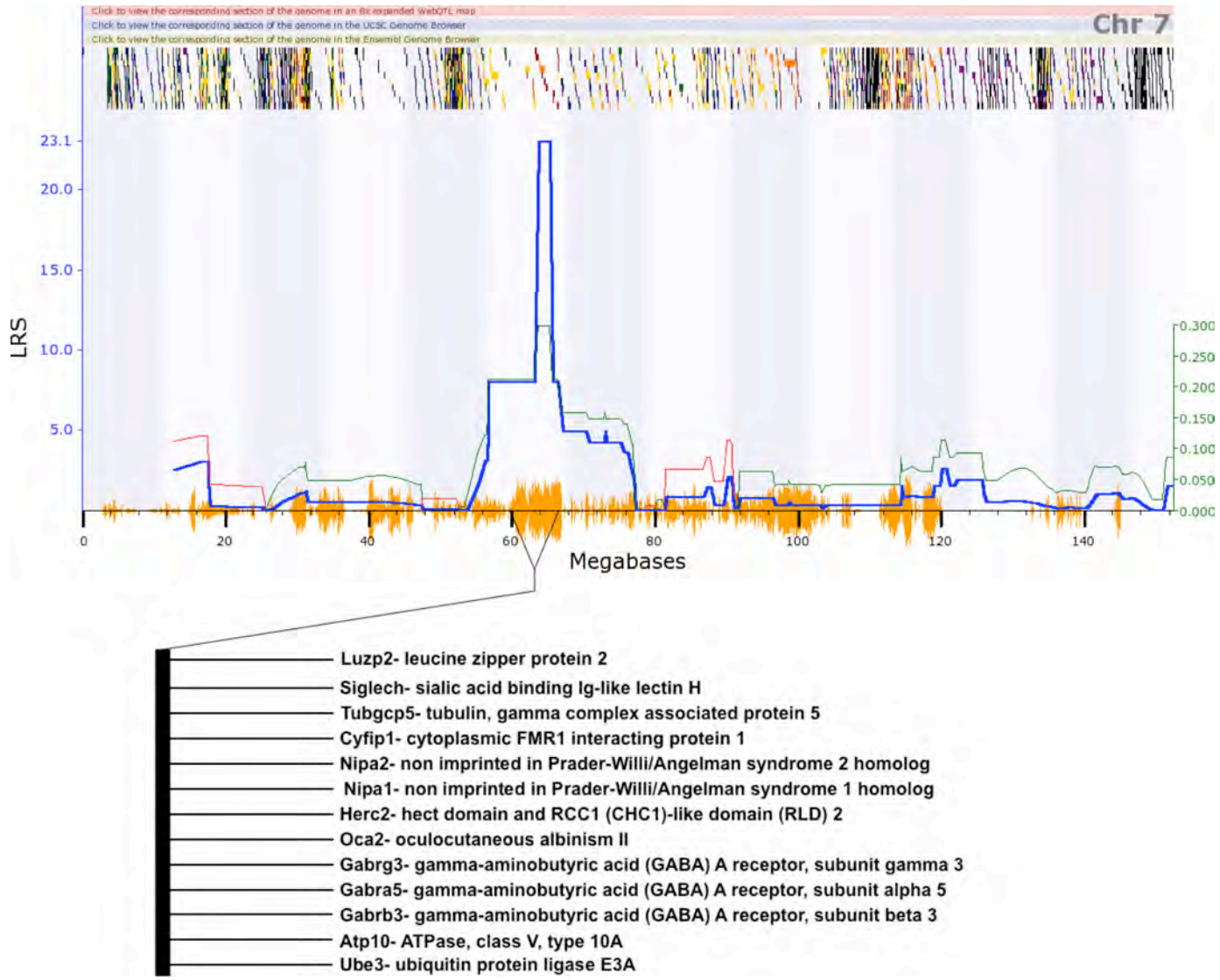


Figure 4. Identity of host genes encoded within the suggestive QTL on chromosome 7. The interval map for chromosome 2 is shown in the top panel. A list of genes encoded within the highly suggestive QTL is shown in the bottom panel. There are no genes within the interval that are known to have a role in immune processes.

We have recently initiated studies designed to determine an ideal timepoint for measuring immune parameters of the lungs of mice following pneumonic Bp1026b challenge. Because susceptible mice often succumb to infection by day 3 post-challenge, we focused on timepoints up to 48 hrs post-infection (12h, 18h, 24h, and 48h). We did not observe a significant difference in immune cell recruitment to the lungs of B6 vs. D2 mice at any of the timepoints. We collected serum and BALF samples from each of the animals and will subject them to multiplex cytokine analysis using a Luminex 32-plex cytokine/chemokine kit in the near future. We are hopeful that

we will identify a number of additional differential phenotypes in the parental strains that can be used to perform QTL analysis. Ultimately, such phenotypes will enable us to model gene networks that are responsible for the differential susceptibility of B6 and D2 mice to *Burkholderia* infection.

B. Key Research Accomplishments

- We have identified a clear readout of differential susceptibility/resistance of B6 and D2 mice following pneumonic infection with Bp1026b.
- We have screened 33 BXD strain strains using survival as a readout for QTL analysis, and we have determined that survival is a complex trait that likely involves genes on chromosomes 4 and 7.
- We have begun our search for additional phenotypes of differential responsiveness of B6 and D2 mice to Bp1026b.

C. Reportable Outcomes

Manuscript submitted for publication

1. *Ceballos-Olvera, I., Sahoo, M., Miller, M.A., del Barrio, L., and Re F.* Inflammasome-dependent pyroptosis and IL-18 protect against *Burkholderia pseudomallei* lung infection while IL-1β is deleterious detrimental.

1.2.2 Identification of cellular pathways responsible for genetic susceptibility to *Burkholderia pseudomallei* infection

A. Introduction

Nod-like receptors (NLR**)** are pattern recognition receptors located in the cytoplasm, which they survey for evidence of danger and infection (1). NLR primarily control activation of the inflammasome**,** a multiprotein complex that contains, in addition to NLR, the adaptor molecule ASC and the protease caspase-1. Activation of caspase-1 in the context of the inflammasome is responsible for the proteolytic processing of the immature form of IL-1β IL-18, a modification required for their secretion and bio-activity. Activation of caspase-1 is also triggering a form of cell death called pyroptosis that effectively restrict intracellular bacteria growth. Both production of IL-1b and IL-18 and induction of pyroptosis have been shown to be protective effector mechanisms against many infectious agents. NLRP3 and NLRC4 are the best characterized NLR molecules. NLRP3 control caspase-1 activation in response to "danger signals", several particles and crystals, and various bacteria, viruses, and fungi. The NLRC4 inflammasome is responsive to a narrower spectrum of activators including cytoplasmically delivered bacterial flagellin and the basal rod constituent of various bacterial Type III secretion systems (T3SS). The T3SS apparatus is used by several bacteria, including *Salmonella, Listeria, Francisella, Shigella, Legionella*, to inject virulence factors into the cytoplasm of target cells.

To identify cellular pathways responsible for genetic susceptibility to *B. pseudomallei* infection we initiated an analysis of mice deficient in the inflammasome. These studies complement the other ongoing analysis of the BXD strains. Using a murine model of melioidosis we have performed a detailed analysis of the role of the inflammasome components NLRP3, NLRC4, Caspase-1 and the effector mechanisms IL-1β, IL-18, and pyroptosis.

B. Key Research Accomplishments

NLRP3 and NLRC4 differentially regulate production of IL-1β and IL-18 and pyroptosis

To identify the pathway responsible for IL-1β and IL-18 secretion in response to infection with *B. pseudomallei*, bone marrow-derived dendritic cells (BMDC) derived from WT mice or mice deficient in the inflammasome componenets ASC, NLRP3, NLRC4, or Caspase-1 were infected in vitro with *B. pseudomallei* and secretion of IL-1β in culture supernatants was measured. As shown in **figure 1A**, secretion of IL-1β was abolished in $ASC^{-/-}$, and $Caspase\text{-}1^{-/-}$ cells. Production of IL-1β during the first four hours of the infection was significantly reduced in $NLRC4^{-/-}$ but not $NLRP3^{-/-}$ cells. However, later in the infection process (8 hrs) $NLRC4^{-/-}$ cells secreted IL-1b at level comparable to WT cells while $NLRP3^{-/-}$ cells showed significantly decreased cytokine secretion. Secretion of IL-18 followed a similar pattern (data not shown). Immunoblotting of the supernatants confirmed processing of IL-1β to the mature 17 kDa form and of caspase-1. The differences in IL-1β and IL-18 secretion were observed regardless of the number of bacteria used to infect cells (m.o.i. 10, 50, or 100, data not shown). Thus, the NLRC4 and NLRP3 inflammasomes are both mediating release of IL-1β and IL-18 by myeloid cell infected with *B. pseudomallei*.

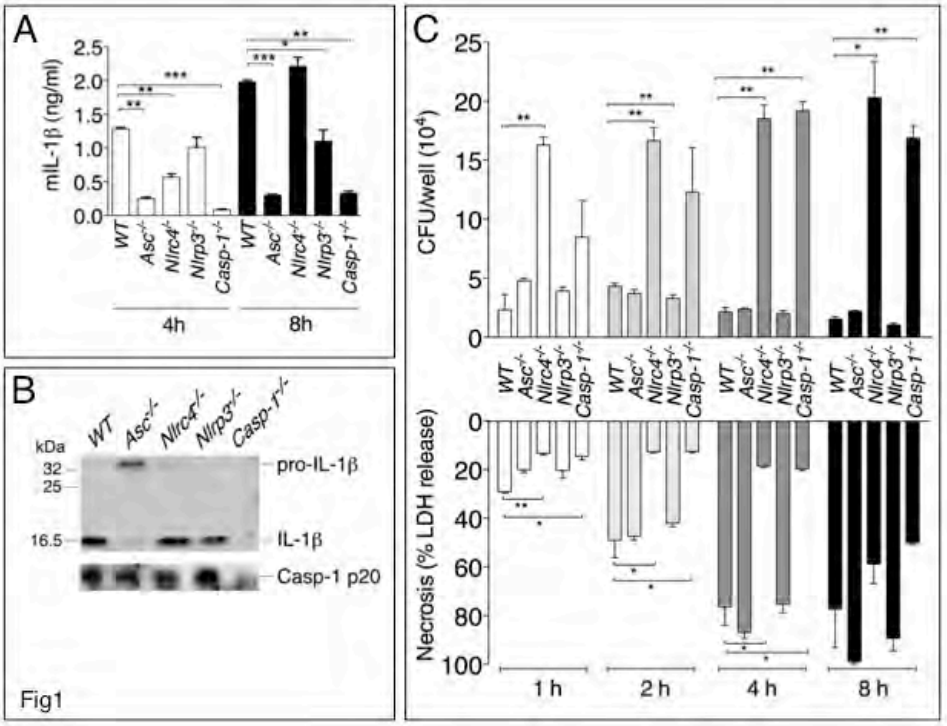


Figure 1. ***NLRP3 and NLRC4 differentially regulate production of IL-1β and IL-18 and pyroptosis*****.** BMDC were infected with *B. pseudomallei* at m.o.i. 10. (A) Secretion of mature IL-1b was measured in conditioned supernatants at the indicated times. (B) Processing of IL-1β and Caspase-1 was detected by immunoblot in 8h conditioned supernatants from A. (C) Cells were lysed and intracellular bacteria growth was quantitated (upper panel). Induction of pyroptosis was measured as LDH release in conditioned supernatants (lower panel). One experiment representative of 4 (A) or three (C) is shown.

Inflammasome-mediated induction of pyroptosis has been demonstrated to be a mechanism to restrict growth of certain intracellular bacteria (2). Induction of pyroptosis and intracellular bacteria proliferation were measured in WT or inflammasome-deficient BMDC infected with *B. pseudomallei*. As shown in **figure 1C**, necrosis of infected cells (as measured by release of LDH in culture supernatants) was significantly reduced in Caspase-$1^{-/-}$ and NLRC$4^{-/-}$ cells compared to WT and NLRP$3^{-/-}$. Importantly, induction of pyroptosis was not lost in cell deficient in ASC, the adaptor molecule that is required for NLRC4 function in response to certain bacteria. NLRC4-mediated pyroptosis induced by other bacteria is also ASC-independent. Consistent with the role of pyroptosis as a mechanism to restrict bacteria growth, considerably less intracellular bacteria were recovered from WT, NLRP$3^{-/-}$ and ASC$^{-/-}$ cells than from Caspase$1^{-/-}$ or NLRC$4^{-/-}$ cells.

Taken together these results show that infection of BMDC with *B. pseudomallei* leads to activation of the NLRC4 and NLRP3 inflammasomes. NLRC4 mediates induction of pyroptosis, restriction of bacteria growth, and IL-1β and IL-18 production during the early phase of the infection. NLRP3 does not control pyroptosis and its role in IL-1β and IL-18 secretion becomes more prominent as the infection progresses. It should be noted that the defective IL-1β production of NLRC$4^{-/-}$ and NLRP$3^{-/-}$ cells cannot be ascribed to the difference in induction of pyroptosis: thus NLRP$3^{-/-}$ cells produce less cytokine than WT cells despite undergoing pyroptosis to the same extent as WT cells. Conversely, NLRC$4^{-/-}$ cells, which are resistant to pyroptosis, still produce less cytokine than WT cells at the early time point.

Role of inflammasomes in murine melioidosis

The role of the inflammasome during *in vivo B. pseudomallei* infection was next analyzed using a mouse model of melioidosis. WT mice or inflammasome deficient mice were infected intranasally with *B. pseudomallei* (100 cfu) and their weight and survival were monitored. All mice started to loose weight 2 days post-infection (not shown). Generally, mice that survived the infection start to recover weight 7 days post-infection. As shown in **figure 2A**, Caspase-1$^{-/-}$, NLRC4$^{-/-}$, and ASC$^{-/-}$ mice were extremely susceptible to melioidosis compared to WT mice. NLRP3$^{-/-}$ mice were also considerably more susceptible than WT mice but slightly more resistant than the other inflammasome deficient mice. Measurement of the bacteria burden in lung, spleen, and liver of infected mice 24 hours (data not shown) and 48 hours post-infection revealed that NLRC4$^{-/-}$ and Caspase-1$^{-/-}$ mice carried considerably higher bacteria in all three organs than WT mice (**figure 2B**). Surprisingly, the bacterial burden of ASC$^{-/-}$ and NLRP3$^{-/-}$ mice was not significantly different from that of WT mice at the tested time points despite their higher mortality.

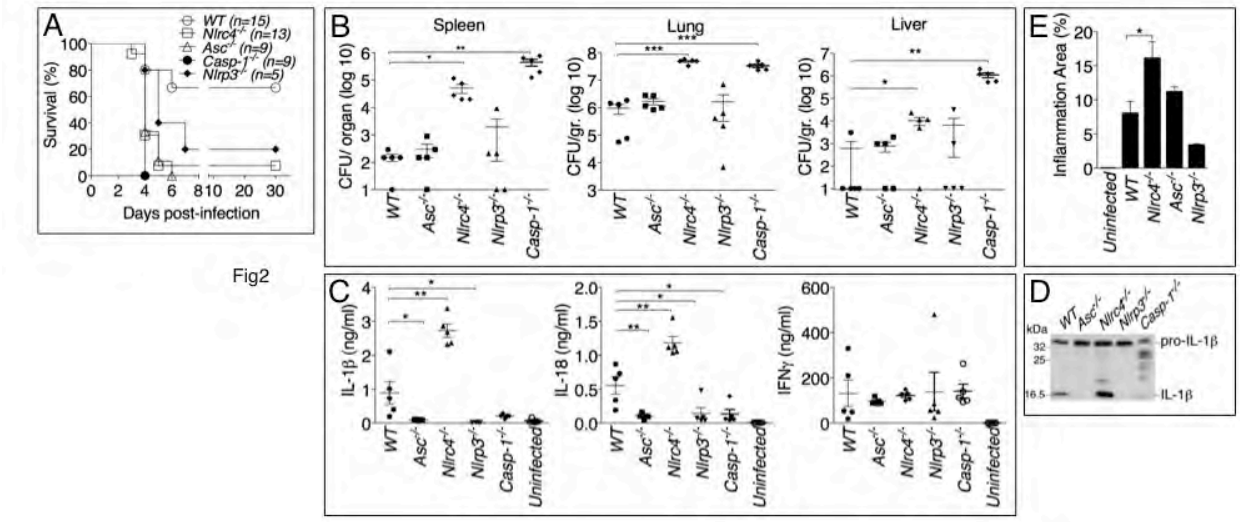


Fig2

Figure 2. Differential contributions of NLRP3 and NLRC4 to melioidosis.
(A) Mice were intranasaly infected with *B. pseudomallei* (100 CFU) and their survival was monitored. (B) Mice were sacrificed 48 hours post-infection and the bacteria burden was measured in organs homogenates. (C) Cytokines were measured in BALF obtained 48 hrs post-infection. (D) Processing of IL-1β was detected by immunoblot in BALF from C. (E) Lung section were stained with H7E and the total area of the inflammatory nodules was measured and expressed as percentage of the total lung lobe area.

It has been demonstrated that activation of the NLRC4 inflammasome by *Legionella* is also dependent on the Naip5 molecule (3), an NLR member that is mutated in the A/J mouse strain. We tested the susceptibility of C57BL/6J-Chr13a/j/NaJ mice, a consomic C57BL/6 strain that carries the A/J Naip5 allele, and found that they were indistinguishable from WT mice (data not shown).

Cytokine levels were measured in bronchio-alveolar lavage fluids (BALF) obtained from infected mice (**figure 2C**). Confirming the in vitro results, IL-1β and IL-18 level were severely reduced in ASC$^{-/-}$, Caspase-1$^{-/-}$ and NLRP3$^{-/-}$ mice. However, quite surprisingly, IL-1β and IL-18 were present in the lungs of NLRC4$^{-/-}$ mice in amount considerably higher than in WT mice. Immunoblotting experiments confirmed that the IL-1β measured by ELISA was in fact the p17 mature form of IL-1β (**figure 2D**). Thus, although the in vitro experiments demonstrated that

both the NLRP3 and the NLRC4 inflammasome contribute to IL-1β and IL-18 production in response to *B. pseudomallei* infection, in vivo it is the NLRP3 inflammasome that primarily mediates production of these cytokines. It is conceivable that the higher bacterial burden due to defective pyroptosis of NLRC4$^{-/-}$ mice is accountable for the higher level of IL-1β and IL-18 in the BALF of these animals. Levels of IL-1α were also significantly higher in NLRC4$^{-/-}$ (not shown). It is interesting to note that the levels of IL-1β and IL-18 in BALF of ASC$^{-/-}$ and Caspase-1$^{-/-}$ mice, although very low, were higher than uninfected mice suggesting the existence of inflammasome-independent mechanisms to produce IL-1β and IL-18, as it has been previously shown for infection with bacteria that cause highly neutrophilic inflammation (4).

Histologic analysis of the infected lungs revealed extensive inflammatory cells infiltration in the lung parenchyma (**figure 2E**). The area of the inflammatory nodules, relative to the total area of the lung lobe, was calculated for each given section and found to be significantly greater in NLRC4$^{-/-}$ and Casp-1$^{-/-}$ mice compared to WT, ASC$^{-/-}$, or NLRP3$^{-/-}$ mice. This result is consistent with the increased level of inflammatory cytokine and chemokines in NLRC4$^{-/-}$ and Casp-1$^{-/-}$ mice.

Taken together these results identified two distinct infammasome-mediated mechanisms that efficiently restrict *B. pseudomallei* growth and pathogenesis: production of the cytokines IL-1β and IL-18 and pyroptosis. The high susceptibility to meliodiosis of NLRP3$^{-/-}$ mice is due to defective cytokines production while that of the NLRC4$^{-/-}$ mice to defective pyroptosis. Caspase-1$^{-/-}$ mice are impaired in both inflammasome effector mechanisms and therefore would be predicted to be more vulnerable to *B. pseudomallei* than mice deficient in ASC or IL-1RI/IL-18 (that lack cytokines but retain pyroptosis) or NLRC4$^{-/-}$ mice (that retain IL-1β /IL-18 production but are deficient in pyroptosis). This prediction turned out to be correct. When mice were infected with only 25 cfu (a non-lethal dose for WT mice) all Casp-1$^{-/-}$ mice rapidly succumbed (not shown). The mean time to death was increased for NLRC4$^{-/-}$ and IL-1RI$^{-/-}$/IL-18$^{-/-}$ double knock-out mice (DKO). Surprisingly, ASC$^{-/-}$ mice, which should be equivalent to DKO, survived longer. This may be explained by the observation that IL-18, although drastically reduced in ASC$^{-/-}$ and NLRP3$^{-/-}$ mice, is still detectable in ASC$^{-/-}$ mice at higher level than uninfected mice (**figure 2C**).

Role of IL-18 and IL-1β in murine melioidosis

We next analyzed the role of the inflammasome-dependent cytokines IL-1β and IL-18 during murine melioidosis. IL-18-deficient mice were extremely susceptible to *B. pseudomallei* infection even when infected with low doses of bacteria that caused no mortality and only mild weight loss in WT mice (**figure 3A**).

In contrast, IL-1RI$^{-/-}$ mice showed increased resistance to *B. pseudomallei* infection compared to WT mice (**figure 3A** and see below). The survival of mice deficient in both IL-18 and IL-1RI (DKO) was indistinguishable from the IL-18$^{-/-}$ mice when the animals were infected with 100 cfu. However, in mice infected with 25 cfu the concomitant absence of IL-18 and IL-1RI provided a significant advantage over IL-18$^{-/-}$ mice ($p<0.05$) suggesting a detrimental role of IL-1RI-mediated signaling in melioidosis (see below).

Confirming the different susceptibility of IL-18$^{-/-}$ and IL-1RI$^{-/-}$ mice, the bacterial burdens in lung spleen and liver of infected IL-18$^{-/-}$ mice were hundred times higher than that of WT mice even at early time points (24 hours post infection, figure 3B). In contrast, significantly lower amount of bacteria were recovered 48 hours post infection from IL-1RI$^{-/-}$ mice compared to WT mice confirming their higher resistance.

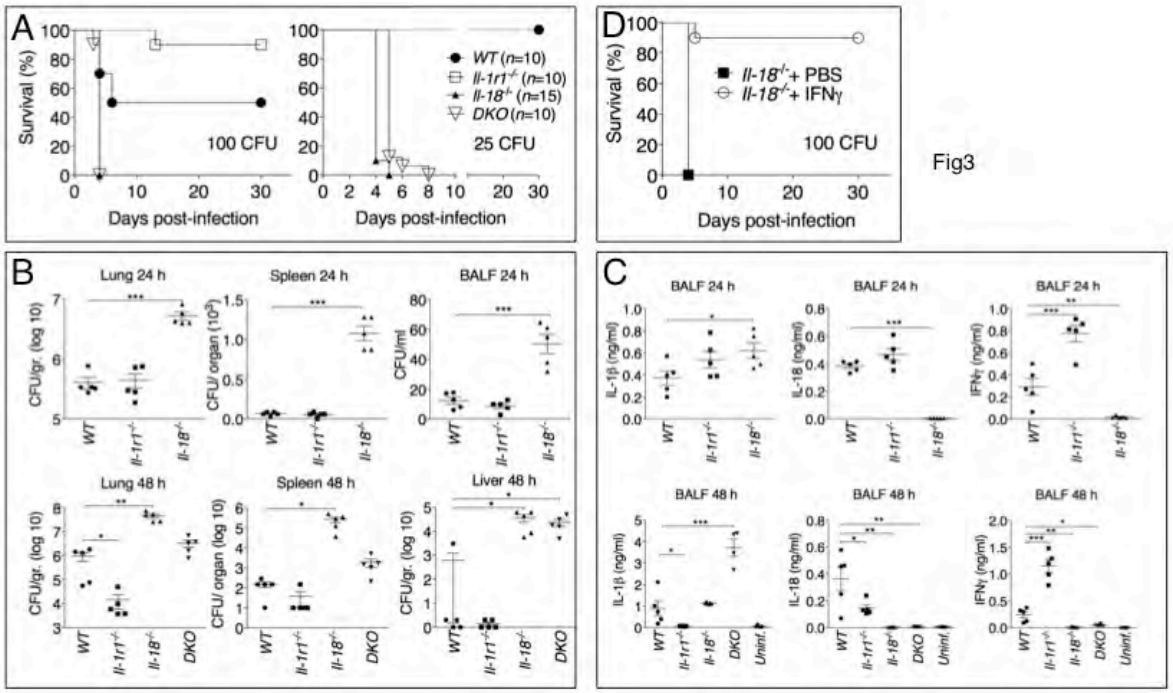


Figure 3. Differential contributions of IL-1 and IL-18 to melioidosis. (A) Mice were intranasaly infected with BP (100 CFU, left or 25 CFU, right) and their survival was monitored. Mice infected with *B. pseudomallei* (100 CFU) were sacrificed 24 hours or 48 hours postinfection and the bacteria burden in organs homogenates (B) or cytokines levels in BALF(C) were measured. (D) IL-18-/- mice were intranasaly infected with *B. pseudomallei* (100 CFU) and their survival monitored. Mice were administered daily injections of PBS or IFNγ (1 μg) for the first 8 days.

Measurement of different cytokines in the BALF obtained from infected mice at 24 and 48 hours post-infection (**figure 3C**) indicated that the levels of IFNγ were drastically reduced in IL-18$^{-/-}$ mice, a finding consistent with the established function of IL-18 as an IFNγ-inducing cytokine. Remarkably, IFNγ level in IL-1RI$^{-/-}$ mice were greatly increased compared to ΩT mice. IL-1β was also significantly increased in DKO mice.

Considering that IFNγ is known to play a protective role during several bacterial infections, including *B. pseudomallei* (5), these results suggest that the reduced resistance to *B. pseudomallei* infection of IL-18$^{-/-}$ mice may be due to lack of IFNγ induction. To test this hypothesis a group of IL-18$^{-/-}$ mice were infected with *B. pseudomallei* and received daily intraperitoneal injections of recombinant IFNγ. As shown in **figure 3D**, exogenous IFNγ completely protected the mice suggesting that IL-18 exerts its protective action primarily through induction of IFNγ.

Deleterious role of IL-1 in melioidosis

The results of **figure 3** showed that IL-1RI$^{-/-}$ mice were more resistant to lung infection with *B. pseudomallei*. This appeared even more evident when mice were infected with higher doses of *B. pseudomallei* that killed all WT mice but only a fraction of the IL-1RI$^{-/-}$ (**figure 4A**). The extent of lung inflammation, as measured by recruitment of neutrophils, macrophages, and dendritic cells, and the number and size of inflammatory nodules, was significantly decreased in IL-1RI$^{-/-}$ mice compared to WT or IL-18$^{-/-}$ mice (**figure 4D-F**). To confirm the notion that IL-1R-mediated

signaling has a deleterious role in our model of melioidosis, WT mice were infected with 100 cfu *B. pseudomallei* and daily administered IL-1β (or PBS) by intraperitoneal injections (**figure 4B**). All mice that received the cytokine succumbed to the infection compared to significantly higher survival of the control group. Injection of IL-1β in non-infected mice had no deleterious effect aside from a transient, negligible weight loss (not shown). The bacteria burden in organs of IL-1β-treated mice 72 hours post infection was dramatically higher than the control group and bacteremia was detected in IL-1β-treated mice but not control mice. Higher number of neutrophils, macrophages and dendritic cells were found in the BALF of IL-1β-injected mice (**figure 4D**). Histological analysis of lung sections showed a dramatic increase in mice treated with IL-1β in the number and size of the foci of infiltrating inflammatory cells and evidence of perivascular edema and airways obstruction.

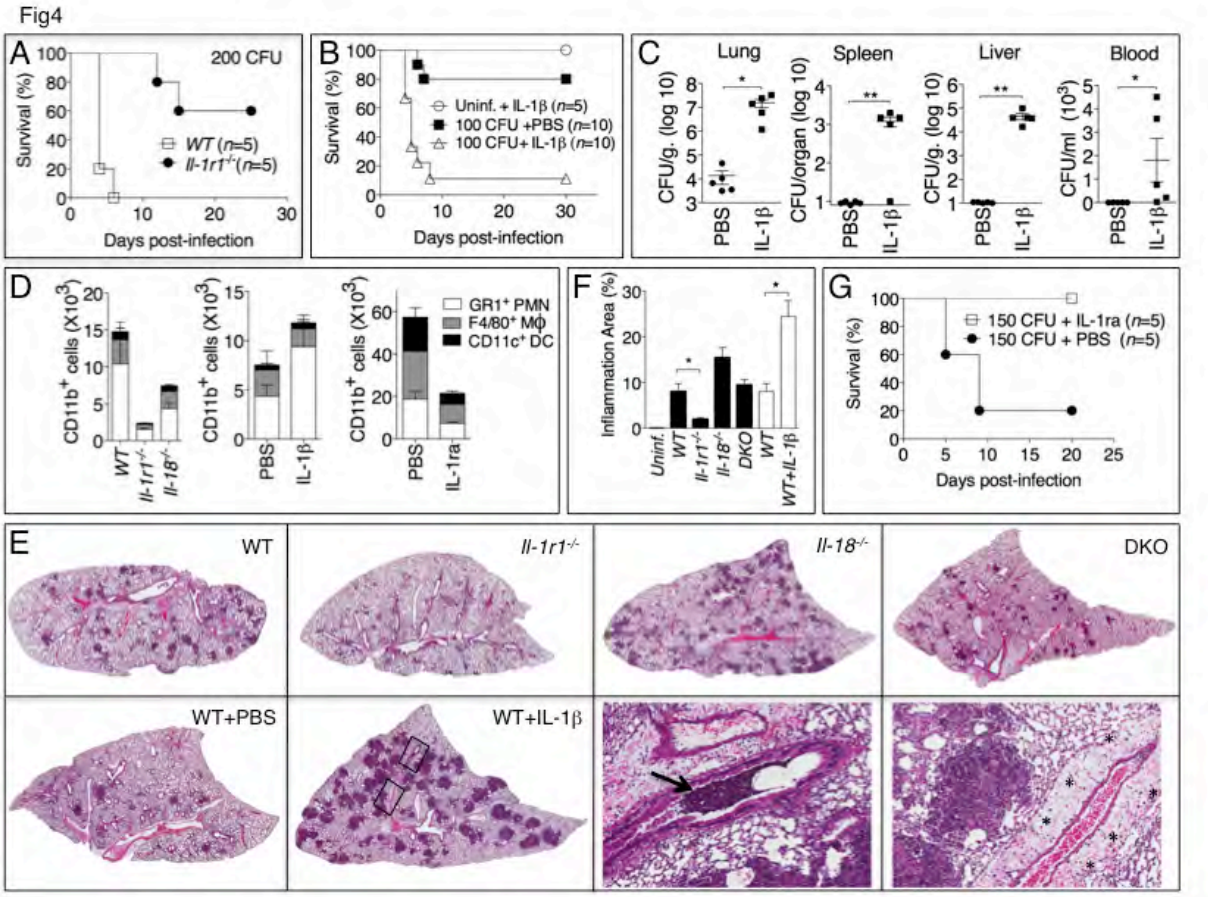


Figure 4. Deleterious role of IL-1β in melioidosis. (A, B) Mice were intranasaly infected with *B. pseudomallei* (200 CFU, A or 100 CFU, B) and their survival was monitored. Mice in B were administered daily i.p. injections of PBS or IL-1β (1μg). One group of mice were treated with IL-1β but not infected. (C) Mice infected and IL-1β -injected as in B were sacrificed 72 h post-infection and the bacterial burden was measured in organs and blood. (D) BALF cells obtained from the indicated infected mouse strains (left) or WT mice injected with IL-1β or IL-1ra were analyzed by cytofluorimetry for myeloid cells composition (neutrophils, macrophages, dendritic cells. (E) Histopathology of lungs of infected mice of indicated genotype at 48 h post-infection (upper row) or WT mice injected with PBS or IL-1β at 72 hrs post-infection (lower row). Bottom right panel show magnification of the indicated insets from WT+IL-1β showing airways obstruction (arrow) and perivascular edema (asterisks).

Additional evidence supporting a deleterious role of IL-1 during melioidosis was obtained by analyzing the effect of administration of IL-1ra, a natural antagonist of IL-1RI. As shown in figure 4G, administration of IL-1ra protected mice from infections with lethal doses of *B. pseudomallei*, decreased lung inflammation (not shown) and recruitment of inflammatory cells (figure 4D).

C. Reportable Outcome

Manuscript submitted for Publication

1. *Ceballos-Olvera, I., Sahoo, M., Miller, M.A., del Barrio, L., and Re F.* Inflammasome-dependent pyroptosis and IL-18 protect against *Burkholderia pseudomallei* lung infection while IL-1β is deleterious detrimental.

D. Conclusions

- In vitro production of IL-1β and IL-18 by BMDC infected with Bp was dependent on NLRC4 and NLRP3 while pyroptosis required only NLRC4.
- Mice deficient in the inflammasome components ASC, caspase-1, NLRC4, and NLRP3, were dramatically more susceptible to lung infection with Bp than WT mice.
- Production of IL-1β and IL-18 was abolished in NLRP3-/- mice.
- The high susceptibility of NLRC4-/- mice was not due to lack of IL-1β and IL-18 but rather to decreased pyroptosis and consequently higher bacteria burden in organs.
- Analysis of IL-1RI- and IL-18-deficient mice infected with Bp revealed that IL-18 is essential for survival primarily because of its ability to induce IFNγ production.
- IL-1RI-deficient mice were more resistant than WT mice to infection. In agreement with a deleterious role of IL-1 during melioidosis, administration of this cytokine increased mortality and bacteria burden in infected mice while administration of IL-1ra had the opposite effect.

In summary our work shows that NLRP3 and NLRC4 play non-redundant roles during *B. pseudomallei* infection by differentially regulating pyroptosis and production of IL-1β and L-18; it reinforces the concept that pyroptosis is an efficient effector mechanism to restrict bacteria growth and dissemination; and it indicates that inhibition of IL-1-mediated signaling may be a beneficial therapeutical approach for the treatment of melioidosis.

E. References

1. Taxman DJ, Huang MT, Ting JP (2010). Inflammasome inhibition as a pathogenic stealth mechanism. Cell Host Microbe. 8:7-11
2. Miao EA, Leaf IA, Treuting PM, Mao DP, Dors M, Sarkar A, Warren SE, Wewers MD, Aderem A. (2010) Caspase-1-induced pyroptosis is an innate immune effector mechanism against intracellular bacteria. Nat Immunol 11:1136-42
3. Lightfield KL, Persson J, Trinidad NJ, Brubaker SW, Kofoed EM, Sauer JD, Dunipace EA, Warren SE, Miao EA, Vance RE. (2011) Differential requirements for NAIP5 in activation of the NLRC4 inflammasome. Infect Immun. 79:1606-14
4. van de Veerdonk FL, Netea MG, Dinarello CA, Joosten LA. (2011) Inflammasome activation and IL-1β and IL-18 processing during infection. Trends Immunol. 32:110-6.
5. Wiersinga WJ, Wieland CW, van der Windt GJ, de Boer A, Florquin S, Dondorp A, Day NP, Peacock SJ, van der Poll T. (2007) Endogenous interleukin-18 improves the early antimicrobial host response in severe melioidosis. Infect Immun. 75:3739-46

1.3. Studies with *Acinetobacter baumannii*

A. Overview

In the quarterly progress report from September 2010, we reported that we were having difficulty establishing pulmonary infection with *Acinetobacter baumannii* (Ab) that resulted in disease symptoms in either of the parental mouse strains. We performed a series of experiments using cyclophosphamide treatment in an effort to reduce the innate response to Ab in hopes of establishing a disease state in one or both of the parental strains. We were unable to establish symptomatic infections using this approach.

In an effort to identify a differential phenotype in the parental strains following pulmonary infection with Ab, we initiated studies to evaluate a series of immune parameters within the lung following intranasal infection with Ab. We performed time-course studies in which mice were infected with 1 x 10^6 CFU of Ab, and then a subset of the mice were sacrificed every 24 hours, bronchiolar lavage (BAL) was performed, and flow cytometric analyses was performed to determine the frequency of neutrophils within the BAL cell population.

The results of these studies have revealed that there is a significant difference in the recruitment of immune cells to the peritoneal cavity of B6 vs. D2 mice 24 hours after intranasal challenge (**Figure 1**). No differences were observed in mice sacrificed at later timepoints (data not shown). Flow cytometric analysis revealed that the bulk of the cells recruited to the lungs of D2 mice were neutrophils (data not shown).

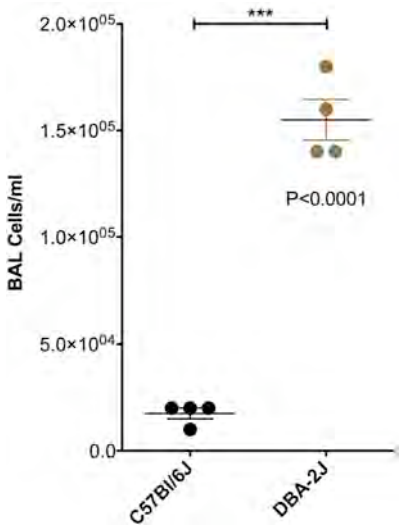


Figure 1. Differential recruitment of immune cells into the lungs following pulmonary infection with *Acinetobacter baumannii*. Mice (4/group) were challenged with either 1 x 10^6 CFU Ab via the intranasal route. The mice were sacrificed 24 hr later and bronchiolar lavage (BAL) was performed. BAL cells were counted using a hemocytometer. Statistical analysis was performed using an unpaired t-test.

We have initiated a series of studies using several lung parameters as phenotypic readouts for BXD analyses. Based on the data shown in Figure 1, recruitment of cell to the lungs is the one of the phenotypes of interest and is the one that we have concentrated most of our efforts on to this point. We have performed screening with 25 BXD strains (all between 10-12 weeks old) and have found that several of the BXD strains are phenotypically similar to D2 mice, some are phenotypically similar to B6 mice, and several strains display intermediate phenotypes (**Figure 2**). Some of the group sizes are too small to yield statistically relevant findings, and additional studies will be performed with these strains as mice of the appropriate age become available. Interval mapping (using GeneNetwork.org) using this cohort of animals reveals a suggestive QTL on chromosome 10 (**Figure 3**). We have established breeding cages for additional strains that will be used to continue this line of investigation and to confirm/narrow the interval on chromosome 10 that appears to correlate with the differential innate immune responses of B6 and D2 mice following Acinetobacter infection.

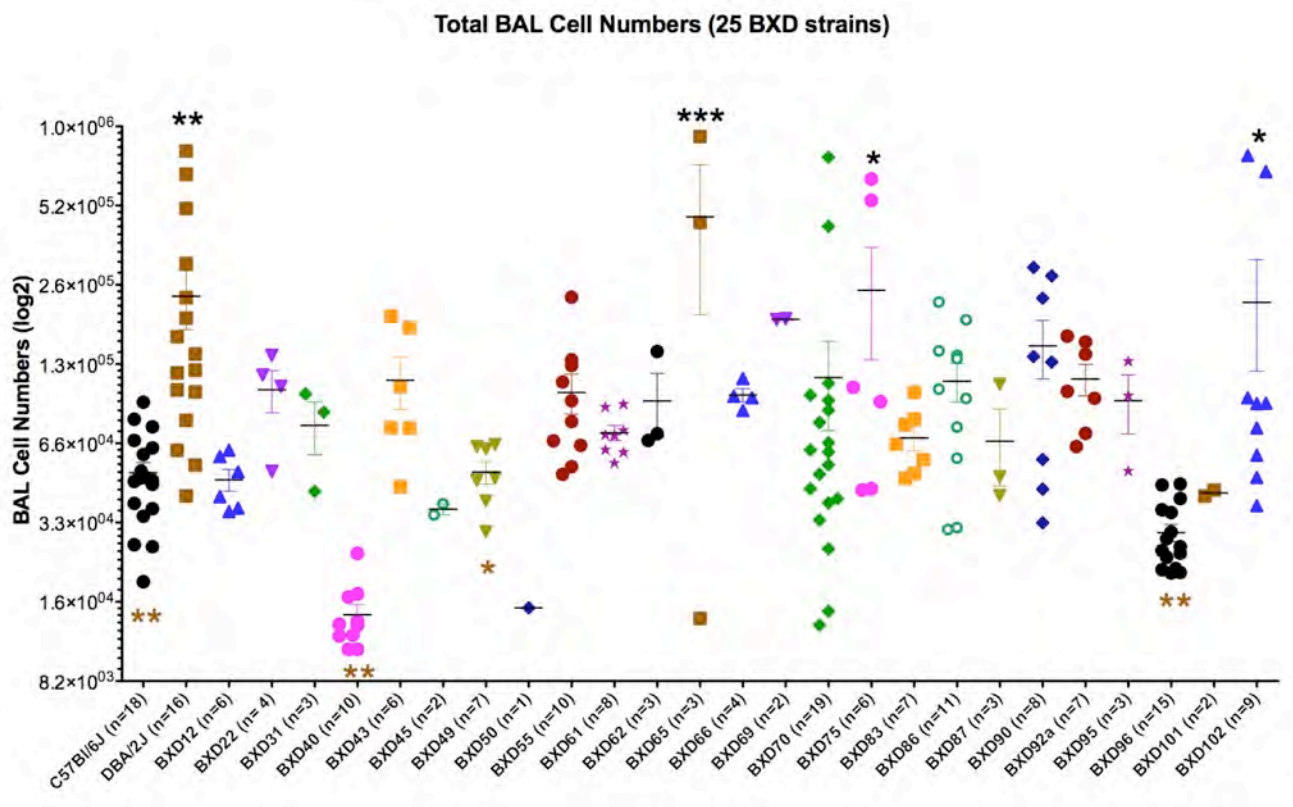


Figure 2. Differential recruitment of immune cells into the lungs of BXD strains following pulmonary infection with *Acinetobacter baumannii*. BXD mice were challenged with 1 x 10^7 CFU Ab via the intranasal route. The mice were sacrificed 24 hr later and bronchiolar lavage (BAL) was performed. BAL cells were counted using a Scepter cell counter. Statistical analysis was performed using one-way ANOVA using Dunnett post tests to compare each of the BXD strains with the C57Bl/6 (statistical

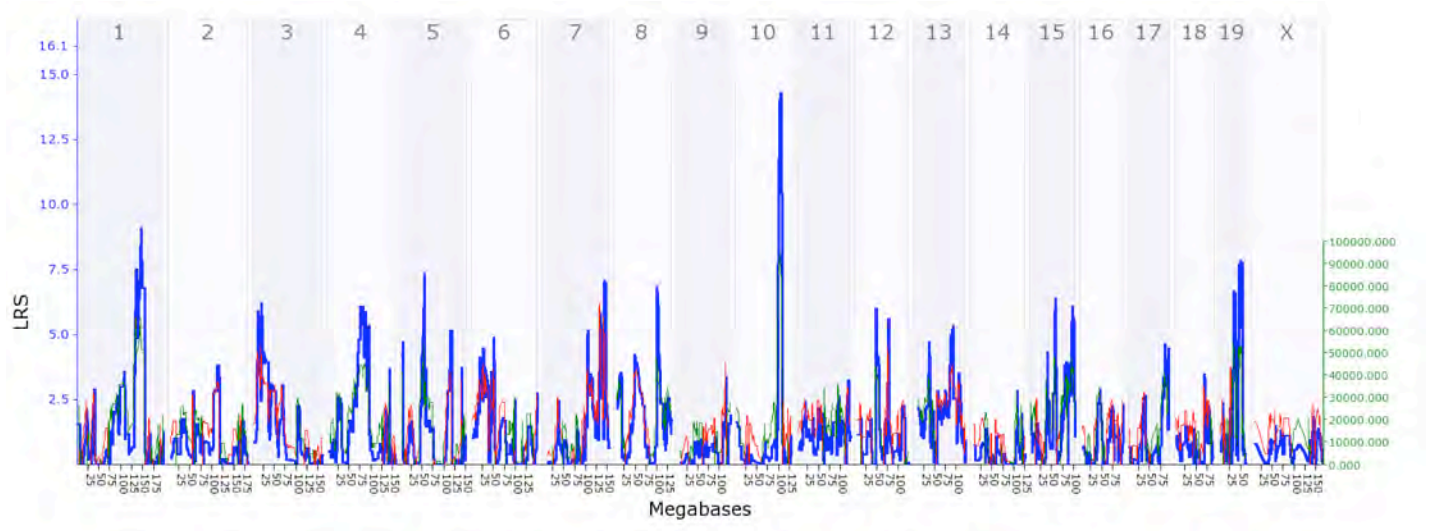


Figure 3. Interval map using recruitment of cells to the lungs of Acinetobacter-infected mice as the query phenotype. A cohort of mice which includes 25 BXD strain were infected intranasally with 1 x 10^7 CFU of Acinetobacter baumannii. Twenty-four hous later, the maice were sacrificed and bronchiolar lavage was performed on each mouse. The number of cells collected from the lungs of each mouse was determined and used as a phenotypic readout for differential innate responsiveness to Acinetobacter. Interval mapping was performed using the in silico resources within GeneNetwork.org. to identify genetic loci that correlate with differential recruitment of immune cells to the lungs of infected mice.

We have also been collecting bronchiolar lavage fluids (BALF) and serum samples upon sacrifice of each of the mice shown in **Figure 2**. Each of the samples will be subjected to multiplex cytokine quantitative analyses using a Luminex 32-plex cytokine/chemokine kit. Approximately half of the BALF samples have already been analyzed and it is apparent that there are significant differences in the production of a number of the analytes by the parental and BXD strains following Ab infection. In BALF samples, we have observed differences in production of 18 of the analytes tested (eotaxin, G-CSF, IFN-g, IL-1a, IL-1b, IL-3, IL-6, IL-17, IP-10, MIP-2, KC, LIF, LIX, MIP-1a, MIP-1b, TNFa, IL-12 p70, and VEGF; **data not shown**). Dr. Cui has performed a preliminary analysis of these data and has identified significant correlation between the levels of a number of these cytokines and has identified three genetic loci that appear to be involved in the differential innate immune responses produced by B6 vs. D2 mice following pneumonic infection with *Acinetobacter* (see modeling section on page 51 below). We have also begun to analyze serum samples collected from this cohort of mice and have observed differences in the levels of 14 of the analytes tested (eotaxin, G-CSF, IFN-g, IL-1a, M-CSF, IL-6, IL-13, KC, LIX, MIP-1a, MIP-1b, RANTES, and TNFa; **data not shown**). In summary, we have identified 33 phenotypes that may be useful in our search for host genetic elements that lead to differential innate immune responsiveness of B6 vs. D2 mice following infection with *Acinetobacter baumannii*.

B. Key Research Accomplishments

- We have identified 33 potentially useful phenotypic differences between B6 and D2 mice 24 hrs after intranasal inoculation with *Acinetobacter baumannii* that will be exploited to identify host genetic loci that are involved in differential innate immune responsiveness.
- We have idenitifed a QTL (on chromosome 10) that potentially correlates with the differential recruitment of immune cells to the lungs in response to pneumonic *Acinetobacter* infection.
- We have determined that a significant correlation exists between the levels of a number of these cytokines measured in the BALF of infected mice and have identified three genetic loci that are potentially involved in the differential innate immune responses produced by B6 vs. D2 mice following pneumonic infection with *Acinetobacter.* Each of these loci include genes that are known to be involved in immune function.

C. Reportable Outcomes

None to date.

D. Conclusions

We have identified a number of clear phenotypic differences between the innate responsiveness of B6 and D2 mice to pulmonary infection with *Acinetobacter baumannii*. We have initiated studies in a cohort of BXD strains and hope to identify host genes that correlate with the differential responsiveness of the parental strains to *Acinetobacter*. Although we have only completed analyses of relatively few BXD strains, we have already identified a significant correlation between several cytokines produced in the lungs and the differential responsiveness of B6 vs. D2 mice to Acinetobacter infection. Moreover, we have preliminarily identified a genetic locus on chromosome 10 that correlates with differential recruitment of immune cells to the lungs and three loci (on chromosomes 6, 14, and 19) that appear to correlate with differential cytokine production following pneumonic infection.

2. Highly Pathogenic Influenza A virus (H5N1) project

2.1 Overview

During the second year of the proposal we continued to assess the susceptibility of several congenic mouse lines and initiated a second phenotypic screen for highly pathogenic H5N1 virus in the BXD mice. This work was done at St Jude Children's Research Hospital in the laboratory of Dr. Richard Webby.

Congenic mouse lines

The B6.D2.11D mouse line was obtained from Drs Davis and Lusis at UCLA and transferred to the animal facility at St Jude Children's Research Hospital. In year 1 we had assessed the susceptibility of this strain to a mouse adapted H1N1 virus (A/PR/8/34) and were preparing to do similar studies using the highly pathogenic H5N1 virus A/Hong Kong/213/03 virus (HK213). B6.D2.11D mice were experimentally inoculated with 10^4 EID_{50} of HK213 virus and morbidity and mortality were monitored (**Table 1**). Compared to the female C57BL/6 control mice, the B6.D2.11D mice lost more bodyweight at the later stages of the infection (day 13), however the difference was not statistically significant. Also, one of the 11D mice succumbed to infection, whereas none of the control mice did. A similar pattern of disease was observed in male mice; the B6.D2.11D male mice lost significantly more weight on day 10 compared to the male C57BL/6 control mice. Also on day 13, the bodyweight of the male B6.D2.11D mice was lower compared to the controls. Although the differences are small and the experiment requires validation, the data are promising and suggest the presence of a gene polymorphism in this locus exacerbating disease severity after H5N1 infection.

Table 1: Weight loss and Mortality in B6.D2.11D mice after inoculation with A/Hong Kong/213/03 virus

		Mortality	Morbidity (% weight loss after infection ± SEM)			
			Day 4	Day 7	Day 10	Day 13
Female	**C57BL/6**	0	94.7 ± 0.6	80.6 ± 0.7	80.1 ± 2.0	94.2 ± 2.8
	B6.D2.11D	25	94.7 ± 1.8	82.0 ± 1.2	79.7 ± 3.8	86.0 ± 4.4
Male	**C57BL/6**	0	94.7 ± 0.7	82.1 ± 1.0	89.4 ± 2.0	98.3 ± 1.4
	B6.D2.11D	0	92.3 ± 1.1	81.3 ± 0.9	80.1 ± 3.9	88.5 ± 4.3

Besides B6.D2.11D, we also obtained two other congenic mouse lines containing *Qivr*'s or QTL for influenza virus resistance on a pure C57BL/6 genetic background; B6.D2.17D and B6.D2.7C containing *Qivr17* and *Qivr7* respectively.

A breeding colony for B6.D2.17D has been established and several groups of mice have been infected with 10^4 or 10^5 EID_{50} of HK213 virus (**Table 2**). Morbidity and mortality after inoculation were monitored. Unfortunately no significant difference in mortality or morbidity was observed in the B6.D2.17D mice compared to C57BL/6 controls.

The B6.D2.7C congenic mice are next on the list to be tested for their susceptibility to H5N1 influenza virus; however poor breeding performance has delayed these experiments considerably. Also our genotyping experiments discovered a change in the genotype and instead of an entire chromosome 7 of DBA/2J genetic origin we obtained mice containing only two-thirds of chromosome 7 of DBA/2J. Fortunately the B6.D2.7D mice as they are referred to now do contain the *Qivr7* we had previously identified. Several homozygous mice have been identified and we are now in the process of expanding the colony.

Table 2: Weight loss and Mortality in B6.D2.17D mice after inoculation with A/Hong Kong/213/03 virus

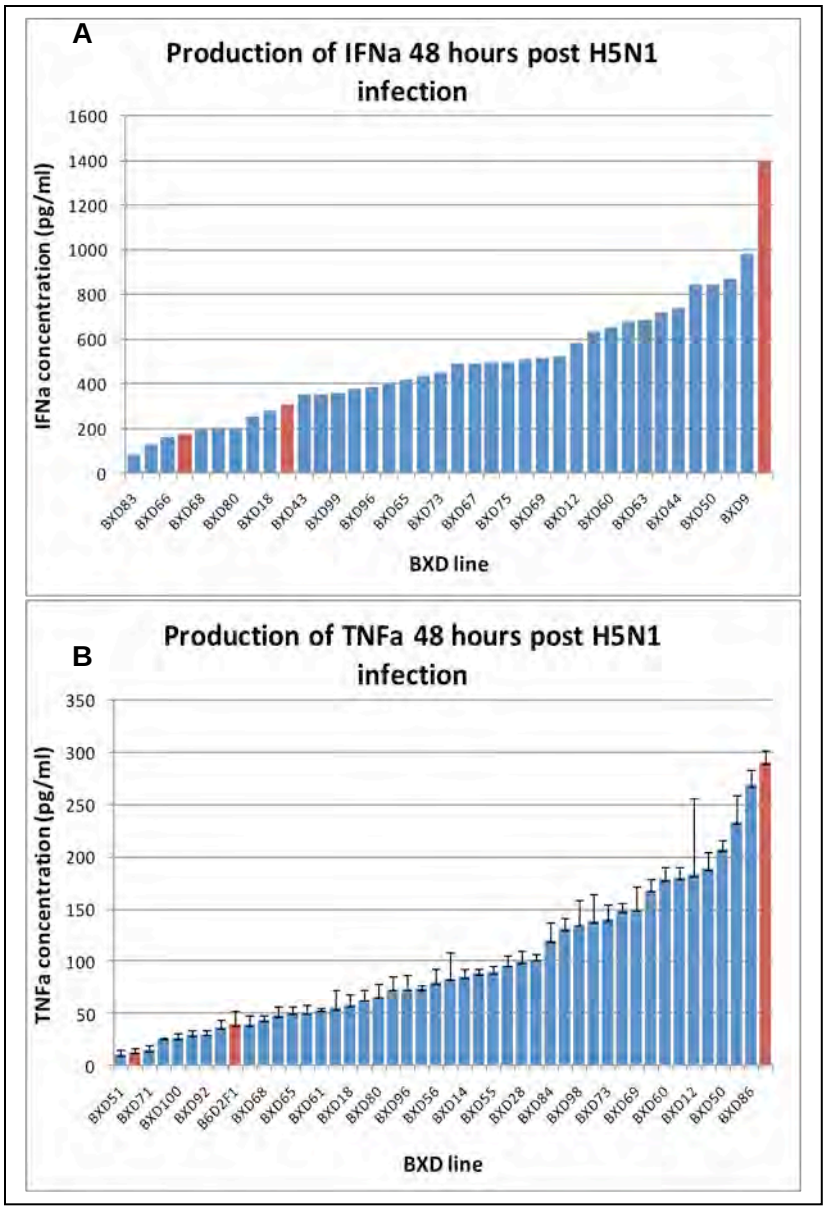
		Dose (EID_{50})	Mortality	Morbidity (% weight loss after infection)		
				Day 4	Day 7	Day 10
Female	**C57BL/6**	10^4	8	96	82	83
	B6.D2.17D	10^4	0	99	84	92
Male	**C57BL/6**	10^5	16	90	78	75
	B6.D2.17D	10^5	50	92	74	71

Novel H5N1 phenotype

While we have worked hard to identify the genes and their polymorphisms in previously identified influenza loci (*Qivr*) associated with death after HK213 virus infection, we have also started testing BXD strains for a second H5N1 phenotype. As reported in the Journal of Virology paper in 2009 from our group (Boon ***et al***), DBA/2J and C57BL/6J mice differ considerably in their response to influenza virus infection with very low LD_{50} values, high influenza titers and excessive production of pro-inflammatory mediators in the DBA/2J mice. High viral titers and production of pro-inflammatory cytokines are hallmarks of H5N1 infection in humans. Therefore we aimed to identify the genetic locus associated with increased production of pro-inflammatory mediators TNFα, CCL2 and IFNα 48 hours after inoculation with H5N1. A total of 43 BXD strains were analyzed for the production of pro-inflammatory mediators and compared to the parental DBA/2 and C57BL/6 mice (**Figure 1**).

Preliminary QTL analysis using the free online QTL analysis tool at www.genenetwork.org, has indicated that a locus on Chromosome 6 is responsible for a significant portion of the observed phenotype for all three inflammatory mediators tested. Two other QTL's, one on chromosome 1 and another on chromosome 12, are considered suggestive. Interestingly, the locus on chromosome 12 contains the *Ahr* gene which is known to differ between C57BL/6 and DBA/2J mice.

Future work will concentrate on identifying the gene on chromosome 6 responsible for the excessive production of pro-inflammatory mediators. Based on additional work in the laboratory, we hypothesize that it is the higher viral load in DBA/2J mice that is initiating this difference in cytokine production and therefore the gene under investigation is likely to influence the replication dynamics of the virus. To validate the result, we have initiated the production of a congenic mouse line containing chromosome 6 of C57BL/6 mice on a DBA/2 background. Compared to the parental DBA/2J mice, the congenic strain will likely produce lower amounts of TNFα, CCL2 and IFNα.



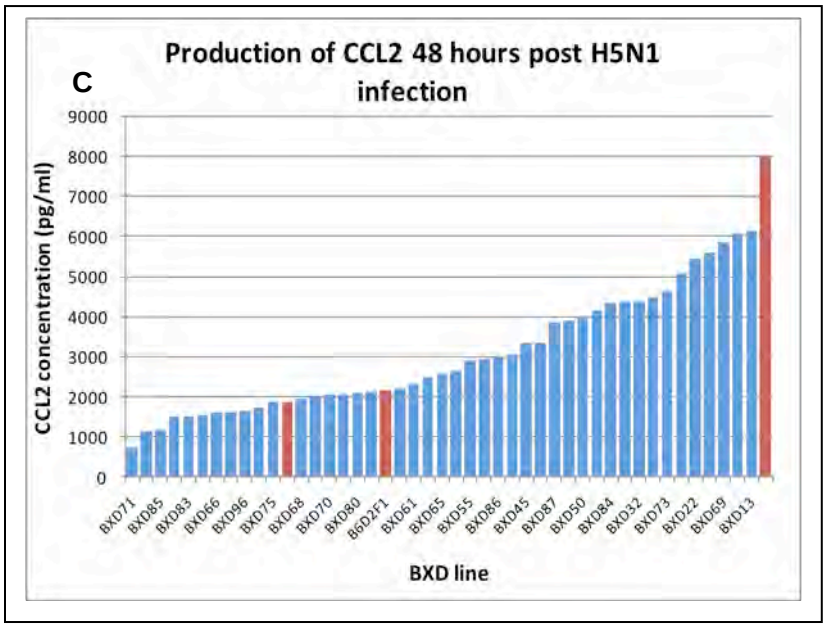


Figure 1: Production of pro-inflammatory cytokines 48 hours post inoculation with Highly pathogenic H5N1 influenza A virus. The production of IFNα[1] (A), TNFα (B) and CCL2 (C) in 43 different BXD lines and the parental DBA/2, C57BL/6 plus the F1 strain. The red bars indicate (from left to right) C57BL/6, B6D2F1 and DBA/2J. [1] Fewer strains were tested for the production of IFNα.

Conclusions

- B6.D2.11D mice are possible more susceptible to H5N1 influenza A virus with an increase in weight loss during the resolution phase of the infection
- B6.D2.17D mice display no obvious phenotype upon inoculation with H5N1 virus.
- DBA/2 produce significantly higher amounts of pro-inflammatory cytokines TNFa, CCL2 and IFNa compared to C57BL/6 mice
- Fourty-three BXD strains have been tested for the production of pro-inflammatory cytokines 48 hours post inoculation with HK213 H5N1 virus
- A significant QTL associated with the early production of pro-inflammatory cytokines has been identified on chromosome 6

Reportable outcomes

Presentations

Lecture at Washington University School of Medicine, St Louis, Missouri.
Lecture at St Louis University, St Louis, Missouri.

Future prospects

Dr. Boon finished his postdoctoral training at St Jude Children's Research Hospital in Memphis and started as an independent investigator at Washington University in St Louis. Although he will be involved in the program, he will not receive any financial support during the 3rd year of the contract

3. *Leishmania major* project

3.1 Overview

In the previous and part of the current funding periods we successfully determined the feasibility of using the BXD resource to study the host response to *Leishmania major* (Lm), a protozoal parasite that causes a disease in mice similar to cutaneous leishmaniasis in humans. In the course of laying this groundwork we determined the optimal parasite dose (1x10^6), host gender (female) and age (7-9 weeks). We also selected a minimal informative set of phenotypes to measure (**Table 1**) and determined the optimum time to take those measurements (3 weeks post-infection).

Table 1

Phenotypes measured	Frequency	Time
clinical score (footpad swelling)	weekly	over 8 weeks
pathogen score (parasite burden in feet)	once	3 weeks
ex vivo cyto/chemokine expression (popliteal LN)	once	3 weeks
ex vivo cyto/chemokine expression (cardiac blood)	once	3 weeks
myelo/lymphoid immunophenotype (popliteal LN)	once	3 weeks
gene expression profile (popliteal LN)	once	3 weeks

The clinical course we observed was consistent with published reports [1, 2], with DBA/2 being more resistant than C57BL/6. In contrast to the differential clinical course, parasite burdens were similar between the BXD parental strains, highlighting distinct underlying mechanisms, and suggesting that differences in the inflammatory response to Lm might be responsible for the contrasting clinical outcomes. In support of this interpretation, draining popliteal lymph node cellularity was significantly higher in C57BL/6 versus DBA/2. To investigate the nature of the differential cellularity exhibited by DBA/2 and C57BL/6 in response to Lm infection, we performed FACS analysis of draining popliteal lymph node cells. We found that the number of MHC class II+, CD19+ B cells was significantly elevated in C57BL/6 versus DBA/2, accounting for the majority of the cellularity difference. This correlation (between enhanced B cell abundance in the draining popliteal lymph node and increased clinical score) has continued to hold through the current funding period as we have begun to score the BXD strains (**Figure 1**) and promises to permit the identification of underlying regulatory QTLs. The correlation between draining popliteal lymph node B cell abundance and clinical score has been noted previously and is thought to relate to the antigen presentation function of B cells [3, 4]. Thus, the hypothesis neutral BXD approach we have employed has succeeded in uncovering a key biological process validated by independent experimental evidence. We anticipate that in the ensuing funding period the BXD approach will permit us to genetically dissect the role of B cells in determining clinical outcome to Lm.

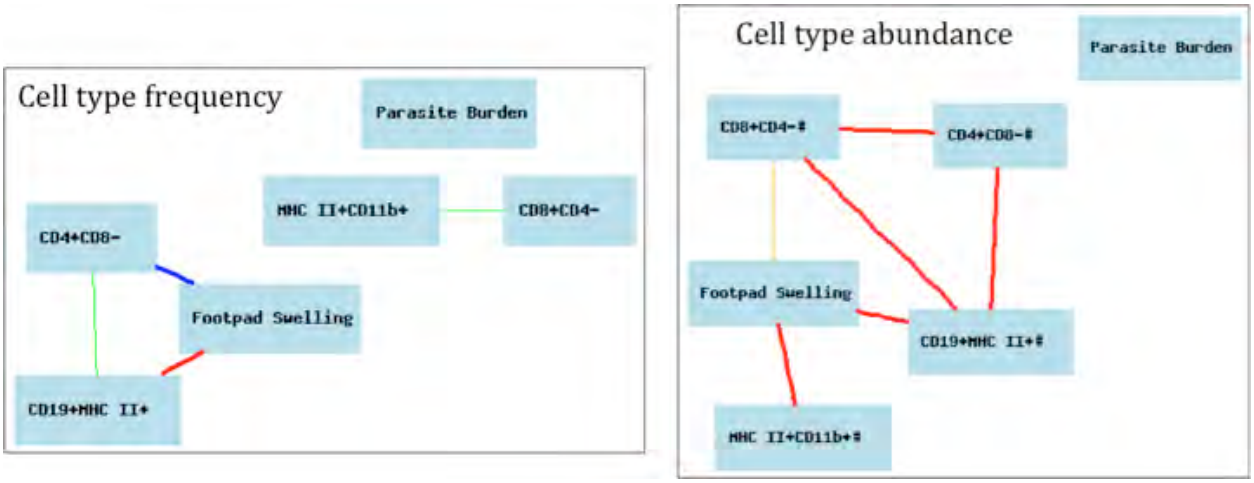
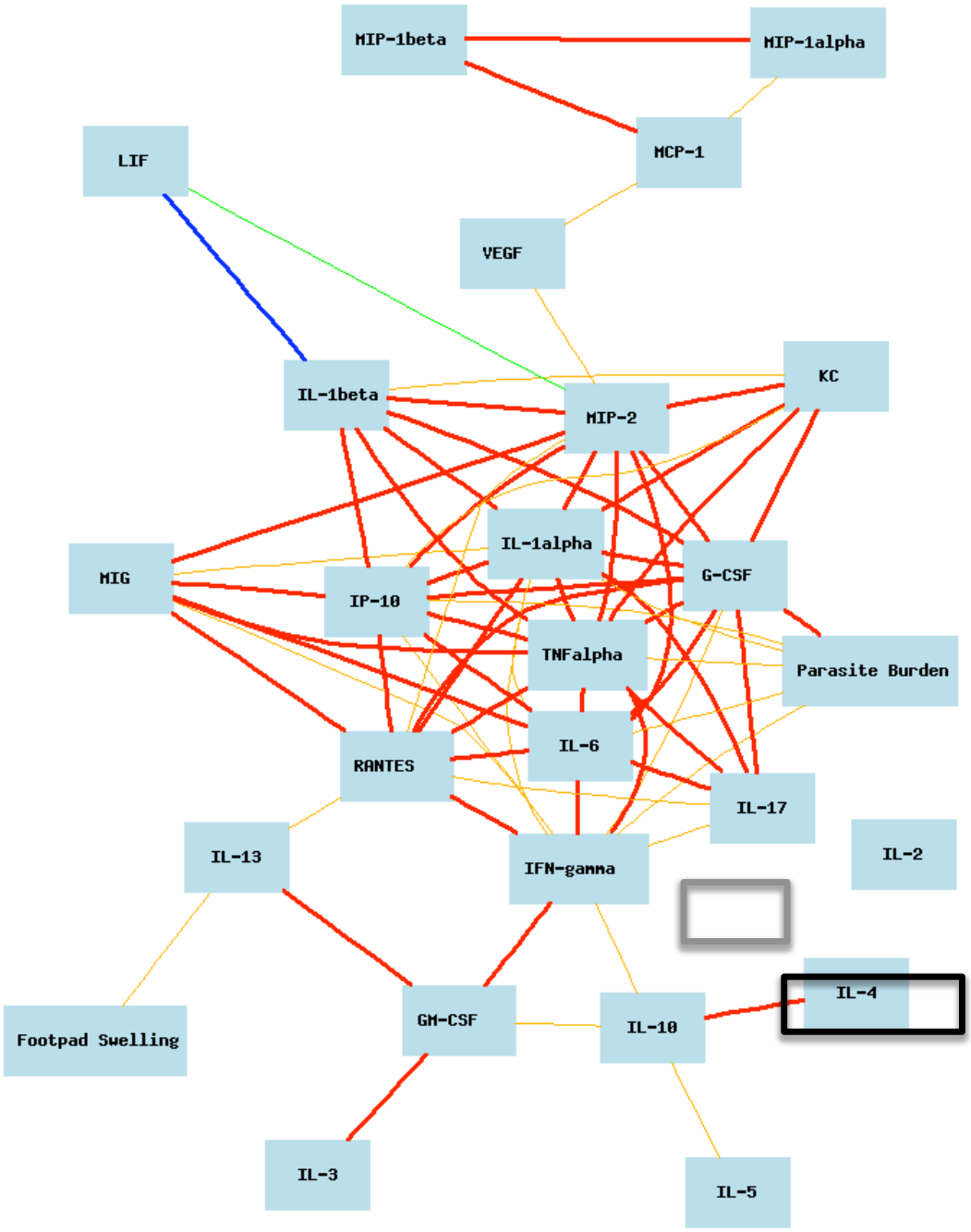


Figure 1. Correlation between CD19+MHC-II+ B-cells and clinical score (footpad swelling) depicted in terms of the frequency and absolute number of the indicated cell types in the popliteal lymph node.

Although, parasite burden did not differ significantly between the parental strains, there was variation among the BXD strains. Thus, we sought to determine whether popliteal lymph node cells exhibit additional differences that could potentially explain the differential parasite burdens exhibited by infected BXD mice. To do this we used a Milliplex bead assay to measure cytokine and chemokine content in supernatants of 48-hour popliteal lymph node cultures. Indeed, GCSF level was found to be strongly correlated with parasite burden (**Figure 2**). As was the case with clinical score and B cell abundance, the correlation between GCSF level and parasite burden has also been noted previously [5, 6]. In this case, evidence suggests that neutrophils mobilized by GCSF to the site of infection ingest parasites and transmit them intact to macrophages within which they replicate and from which they eventually spread. Thus, the hypothesis neutral BXD approach has succeeded in uncovering another key biological process that has been validated by independent experimental evidence. Furthermore, analysis of the data in Gene Network reveals 2 suggestive loci regulating parasite burden (**Figure 3**). Although neither QTL interval contains the gene for GCSF, we hypothesize that the activities that map to each QTL will act in the GCSF/neutrophil/macrophage axis.

Figure 2. Correlation between GCSF level and parasite burden.



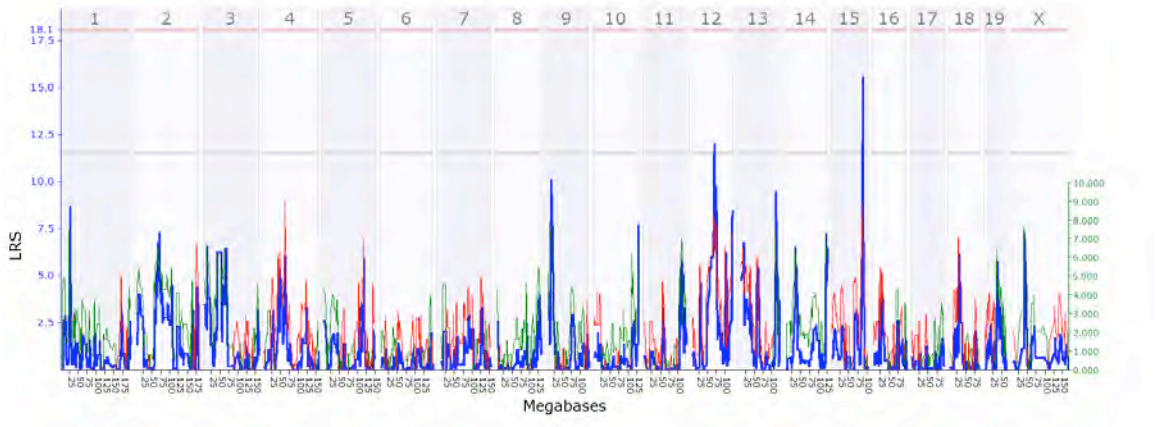


Figure 3. QTLs regulating parasite burden. Note suggestive QTLs on chromosomes 12 and 15.

3.2 Key Research Accomplishments

1. Identified 2 suggestive loci regulating parasite burden.
2. Identified one strong statistical correlation between ex vivo popliteal lymph node GCSF expression level and parasite burden.
3. Identified one strong statistical correlation between popliteal lymph node B cell abundance and footpad swelling.
4. Based on (2) and (3) above, we have determined that assessment of key phenotypes on even as few as 30 BXD mice is sufficient to rapidly triangulate essential host immune features underlying clinical presentation and pathogen score (draining lymph node B cell abundance and GCSF level, respectively). The causal connections underlying statistical correlations unearthed in our analysis have been elucidated experimentally in multiple independent scientific reports REF. Thus, our study serves as a proof of principle demonstration of the power and utility of the BXD resource (combined with appropriate phenotypic measures) to rapidly identify key features of the host response to a given infectious pathogen. An added benefit is the concomitant identification of genetic loci that regulate the key features.

1.3 Reportable Outcomes

Based on year 1 pilot studies with BXD parental strains (C57BL/6 and DBA/2) that established optimal parasite infection dose, host age and gender, phenotypes and measurement intervals, we have collected phenotypic data from an initial cohort of *Lm* infected BXD strains. To date, we have scored 30 BXDs for footpad swelling, 24 for parasite burden, 23 for popliteal lymph node cyto/chemokine expression, 5 for popliteal lymph node immunophenotype.

3.4 Conclusions

Our year 1 studies with BXD parental strains (C57BL/6 and DBA/2) established that the BXD mouse resource would be well suited for the analysis of the host response to *Lm*. Our current year 2 studies in which we have scored an initial cohort of BXD strains fulfill this prediction and motivate us in the upcoming year 3 to score additional BXD strains up to 70 in total. Though currently based on the analysis of up to 30 BXD strains (1-5 mice/strain), our analysis has sufficed to reveal two key correlations that previous more conventional studies have shown to

be causal REF. Our year 2 data indicate that the clinical score (footpad swelling) is strongly correlated with B cell abundance in the draining popliteal lymph nodes. Further, they show that parasite burden is strongly correlated with popliteal lymph node GCSF level. Thus, our preliminary data reveal the power of the BXD approach to rapidly focus attention on key immunological parameters controlling the host response to a given infectious pathogen. In the 3rd year of this project, in addition to bringing our analysis up to 70 BXD strains, we will add expression profiling of RNA collected from draining popliteal lymph nodes. This data will permit detection of eQTLs that regulate the host response to *Lm*. Further by the application of Bayesian analysis, it may be possible to discern cause/effect relationships between specific DNA variants, expression and phenotype QTLs.

3.5 References

1. Reiner, S.L. and R.M. Locksley, *The regulation of immunity to Leishmania major.* Annu Rev Immunol, 1995. **13**: p. 151-77.
2. Baldwin, T.M., et al., *The site of Leishmania major infection determines disease severity and immune responses.* Infect Immun, 2003. **71**(12): p. 6830-4.
3. Smelt, S.C., et al., *B cell-deficient mice are highly resistant to Leishmania donovani infection, but develop neutrophil-mediated tissue pathology.* J Immunol, 2000. **164**(7): p. 3681-8.
4. Ronet, C., et al., *Leishmania major-specific B cells are necessary for Th2 cell development and susceptibility to L. major LV39 in BALB/c mice.* J Immunol, 2008. **180**(7): p. 4825-35.
5. Tacchini-Cottier, F., et al., *An immunomodulatory function for neutrophils during the induction of a CD4+ Th2 response in BALB/c mice infected with Leishmania major.* J Immunol, 2000. **165**(5): p. 2628-36.
6. Kohler, A., et al., *G-CSF-mediated thrombopoietin release triggers neutrophil motility and mobilization from bone marrow via induction of Cxcr2 ligands.* Blood, 2011. **117**(16): p. 4349-57.

4. Severe Acute Respiratory Syndrome (SARS co-V) project

We have grown up the wild-type SARS-CoV (Urbani strain) stocks and have their infectious titers determined. However, we encountered some difficulty in obtaining high-titer stocks for the mouse-adapted MA-15 virus – in the first couple of attempts the MA-15 virus only grew to titers in the 10^5 $TCID_{50}$/ml range in Vero-E6 cells. We thus spent some time optimizing the conditions for propagating the MA-15 virus. We have just worked out the conditions for generating MA-15 virus $\geq 2 \times 10^6$ $TCID_{50}$/ml and are now scaling up the culture to produce large quantity of virus stocks for the animal experiments.

Reportable Outcomes
None

5. Mouse Genomics Core

5.1 Collection of spleen gene expression data from BXD and inbreed strains

This research program relies on massive gene expression data for the spleen from the BXD strains and from common inbred strains. These data are of tremendous use for cloning candidate genes and defining gene networks that modulate host-pathogen interactions and the lethality of select agent. During this past period, we have completed data collection for an extremely large spleen gene expression analysis. Spleens of untreated young adult strains was profiled using the Affymetrix GeneChip Mouse Gene 1.0 ST array that contains approximately 34,728 probe sets and that measures the expression of approximately 29,000 well defined transcripts and essentially all known protein coding genes. This array is an "exon style" array with multiple probes in all known exons of each gene (an average of about 27 probes per gene). A total of 228 chips from 109 strains, including 80 BXD strains, both parental strains (C57BL/6J and DBA/2J), and both reciprocal F1 hybrids (B6D2F1 and D2B6F1), and 25 other common inbred strains were quantified. In most cases, two arrays were processed per strain—one for males and one for females (**Table 1**). This data set will be used for transcriptome-mapping and candidate gene analysis by our own group but also by the entire research community. The data are completely open access and are accessible at www.genenetwork.org.

Table 1. Spleen expression data by strain and numbers per strain

BXD strain	Number	BXD strain	Number	Inbreed	Number
BXD1	2	BXD56	2	C57BL/6ByJ	2
BXD2	2	BXD60	2	C57BL/6J	4
BXD5	2	BXD61	2	C57BL/10J	2
BXD6	2	BXD62	1	C57BLKS/J	2
BXD8	3	BXD63	2	CBA/CaJ	2
BXD9	2	BXD64	3	D2B6F1	2
BXD11	2	BXD65	2	DBA/2J	1
BXD12	3	BXD66	3	FVB/NJ	3
BXD13	1	BXD67	2	ILS	2
BXD14	2	BXD68	2	ISS	1
BXD15	3	BXD69	3	KK/HlJ	3
BXD16	3	BXD70	4	LG/J	2
BXD18	2	BXD71	2	MOLF/EiJ	3
BXD19	3	BXD73	2	NOD/LtJ	2
BXD20	2	BXD74	2	NZB/BiNJ	3
BXD21	2	BXD75	3	NZO/HlLtJ	1
BXD22	2	BXD77	3	NZW/LacJ	2
BXD23	1	BXD78	2	PL/J	1
BXD24	2	BXD79	2	PWD/PhJ	1
BXD25	4	BXD80	2	PWK/PHJ	1
BXD27	2	BXD81	2	SJL/J	2
BXD28	3	BXD82	1	WSB/EiJ	2
BXD29	1	BXD83	3	129P3/J	2
BXD31	2	BXD84	3	129S1/SvImJ	2
BXD32	2	BXD85	2	129X1/SvJ	2
BXD33	2	BXD86	2	B6D2F1	3
BXD34	2	BXD87	2	BALB/cByJ	2
BXD36	2	BXD89	1	BALB/cJ	2

BXD strain	Number	BXD strain	Number	Inbreed	Number
BXD38	2	BXD90	2	BTBR T+ tf/J	1
BXD39	2	BXD92	2		
BXD40	2	BXD93	1		
BXD42	2	BXD95	2		
BXD43	1	BXD96	2		
BXD44	2	BXD97	3		
BXD45	2	BXD98	2		
BXD48	1	BXD99	2		
BXD49	3	BXD100	2		
BXD50	1	BXD101	2		
BXD51	3	BXD102	2		
BXD55	2	BXD103	2		

5.2 BXD strains colony for DoD select agents

We maintained more than 500 cages for DoD including most of JAX BXD strains and all of the UTHSC strains. During 2010–2011, we have provided more than 50 BXD strains to the research program and have helped to set up new breeding cages in the RBL area for multiple DoD projects. Moreover, we are now in the process of developing ~80 new BXD strains in order to increase the power of genetic analysis. Of these new strains, ~ 7 new strains have passed the 8^{th} generation of inbreeding, and ~50 strains have passed the 6^{th} generation.

5.3 Sequence of DBA/2J

Sequence polymorphisms between B6 and D2 are a key resource for QTL analysis and for eQTL cloning. We have now sequenced the D2 genome using short read shotgun sequencing. We have produced ~ 8.26 billion paired-end and mate-pair reads that range in length from 25 to 100 nt using both Illumina and ABI SOLiD platforms. We have used six independently generated libraries. Of these reads, 4.49 billion (261.98 Gb nucleotides) have been mapped to the reference genome of C57BL/6J using the MAQ program and Corona Lite for Illumina and SOLiD reads, respectively. Overall, we have generated approximately 100-fold coverage of the DBA/2J genome (**Table 2**) and sequenced 99.96% of the DBA/2J genome (excluding gaps in the reference genome). Over 99.66% and 95.05% of the reference genome is covered at a depth of at least 10X and 50X, respectively.

Table 2. Summary of sequencing data for the DBA/2J genome

Sequencing platform	Clone insert size	# of total reads (Billion)	# of total nucleotides (Gb)	# of mappable reads (Billion)†	# of mappable nucleotides (Gb)†	Sequence depth (X)*
ABI SOLiD	1-2 kb	1.29	32.27	0.80/1.23	19.95/30.83	7.67 /11.86
	2-3 kb	4.32	163.89	1.92/3.09	72.39/114.87	27.84/44.18
	3-4 kb	1.05	46.06	0.34/0.52	16.28/24.98	6.26/9.61
Subtotal		6.66	242.22	3.06/4.84	108.62/170.23	41.77/65.47
Illumina GA2	101 bp	0.38	38.41	0.35	35.15	13.52
Illumina HiSeq	101 bp	1.22	123.01	1.08	118.21	45.46
subtotal		1.60	161.42	1.43	153.36	58.98
Total		8.26	403.64	4.49/6.27	261.98/323.59	100.75/124.45

*We take the C57BL/6J reference genome size be 2.6 Gb. Read depth was determined through the calculation of all mappable nucleotides divided by the size of C57BL/6J reference genome.

† The number of mappable reads and nucleotides for data from ABI SOLiD using ABI Corona Lite (top) and MAQ (bottom) mapping tools.

For the Illumina HiSeq data, we employed 'amplification-free Illumina sequencing-library' to improve the mapping and assembly of G+C biased genome

We have detected 4,875,300 SNPs variants between the DBA/2J and the C57BL/6J reference genomes at a high confidence threshold, including 4,160,570 and 4,090,000 SNPs using SOLiD (Corona Lite) and from GA2 (MAQ), respectively. Of these, 3,375,200 SNPs are shared between platforms (**Figure 1**). 1.3 million SNPs were detected exclusively using the SOLiD platform. We randomly selected and resequenced 262 SNPs not found in dbSNP, including 128 detected exclusively using the SOLiD platform and 134 detected exclusively using the Illumina platform. The false positive rate was 2.34% and 3.73% for SNPs identified uniquely by SOLiD and Illumina and, respectively. Assuming that all 3.375 million SNPs detected by both platforms are valid, the false positive rate decreases to around 0.44% and 0.65% SOLiD and Illumina, respectively. In summary, the genomes of DBA/2J and C57BL/6J differ from each other at 4.830 million SNPs.

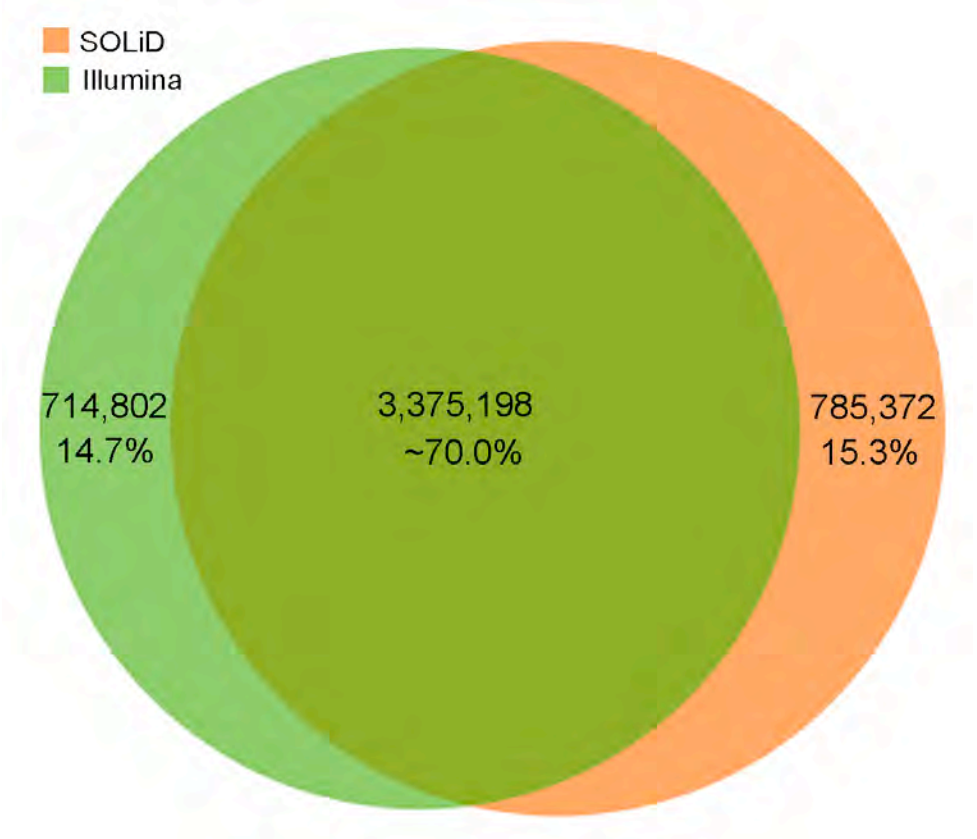


Figure 1. Comparison of SNPs from SOLiD and Illumina platforms. We observe similar numbers of SNPs that are only called by one platform, suggesting that both platforms can complement the SNP detection.

Although SNPs are the most prevalent type of polymorphisms present in the genome, indels are also a major source of genetic variations. To date, comparatively few indels (~1,000) have been identified among between the B6 and D2 (http://www.informatics.jax.org/). Although it remains challenging to detect the indels, we combined MAQ and Pindel and were able to detect up to 100 kb deletions and up to 84 bp insertions. We detected a total of 686,243 small indels including 339,614 insertions, 346,629 deletions (**Figure 2**). SOLiD small indel pipeline detected 121,791 deletions up to 11 bp and 96,973 insertions up to 3 bp. MAQ only detected 9,049 deletions up to 32 bp and 5,700 insertions up to 19 bp. By contrast, Pindel detected 279,115 deletions up to 32,759 bp and 283,375 insertions up to 84 bp. Approximate 16.22% deletions and 13.00% insertions are detected by SOLiD and Pindel for Illumina.

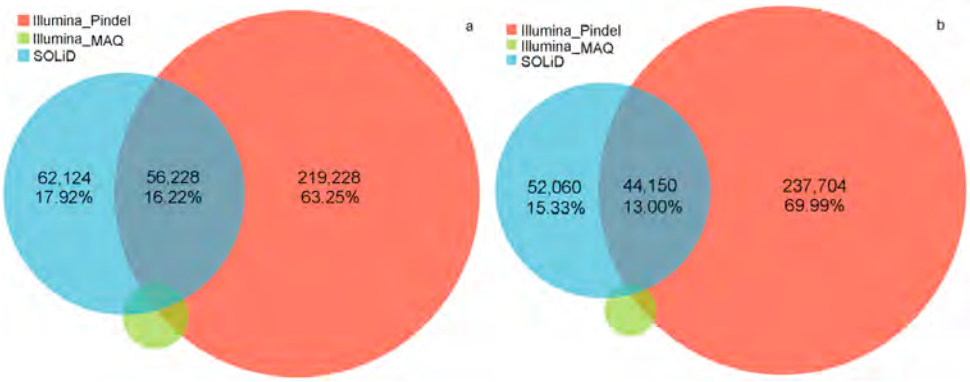


Figure 2. Indels detected by Pindel, MAQ and SOLiD small indel pipeline for Illumina and SOLiD platforms.

We also detected deletion and copy number variance (CNV) except SNA and Indel. We used inter-reads approaches to detect the deletions larger than 100 bp. To detect the full range of sizes of these large indels in the mouse genome, we have constructed four independent libraries with different insert sizes, including 200bp paired-end for Illumina, 1k-2k, 2k-3k and 3k-4k mate-pair for SOLiD. We detected 9,347 deletions for Illumina using the BreakDancer software, and 9,201 large deletions using the ABI large indel pipeline. These indels range in size up to 98,202bp and 970,319 bp for SOLiD and GA2, respectively. We used the EWT method to detect CNV. We observed 16,304 CNVs, including 4,284 gains and 12,020 losses (average length of 34.88 kb and 51.74 kb, respectively).

5.4 Reportable outcomes

Publications

1. Wang X, Agarwala R, Capra JA, Chen Z, Church DM, Ciobanu DC, Li Z, **Lu L,** Mozhui K, Megan K Mulligan MK, Nelson SF, Pollard KS, Taylor WL, Thomason DB, Williams RW: High-throughput sequencing of the DBA/2J mouse genome. **BMC Bioinformatics** 2010, 11(Suppl 4):O7 (see appendices)
2. Abstract:
   Wang X, Mulligan MK, Mozhui K, Lu L, Chen Z, Nelson SF, Taylor WL, and Williams RW: Comparison of variant detections using whole genome sequencing of the DBA/2J mouse strain. The 10th Annual Complex Trait Consortium (CTC) Meeting, Washington DC, USA, June 22~25, 2011 (see appendices).

6. Construction of gene network models

6.1 Introduction

The C57BL/6J (B6) and DBA/2J (D2) strains of mice have a differential susceptibility (difference in 100% lethal dose > 1000,000 fold) to infection with *Chlamydia psittaci*. Previously, a QTL on chromosome 11 (*Ctrq3*), containing two polymorphic innate immune genes (*Irgm2* and *Irgb10*) in the family of immunity-related GTPases (IRGs), was identified as being responsible for the variation in susceptibility to *C. psittaci* in BXD mice. We developed and implemented a Bayesian network modeling method to connect the QTL with immune parameters and disease status in the BXD mice. Several predictions of the Bayesian model were validated through subsequent experiments. This work serves as a proof of concept for a systems genetics approach using BXD mice to study infectious diseases. We have begun to analyze the cytokine levels in the BALF and serum of BXD strains infected with *Acinetobacter baumannii* in order to explain the genetic basis of disease outcome after infection with this high priority pathogen.

6.2 Body

BXD strains with the D2 genotype at *Ctrq3* locus are more susceptible to infection with *Chlamydia psittaci* than strains with the B6 genotype; however, the specific molecular mechanism through which the genotype impacts disease status is unknown (7). To elucidate the pathways that connect *Ctrq3* genotype with disease outcome, we created a Bayesian network model from three sets of data: (1) *Ctrq3* genotype, (2) intermediate immune-related phenotypes, including quantitative variables associated with cytokines, macrophages, neutrophils, and pathogen load, and (3) disease status as quantified by the ratio of weight 6 days after infection to weight at the time of infection. Bayesian networks are graphical models that represent the dependence structure among multiple interacting variables. They consist of variables (nodes) that are connected by directed edges representing conditional dependencies between the nodes. The relationships implied by the network provide testable hypotheses that were used to assess the validity of the model and investigate the relationships in the network.

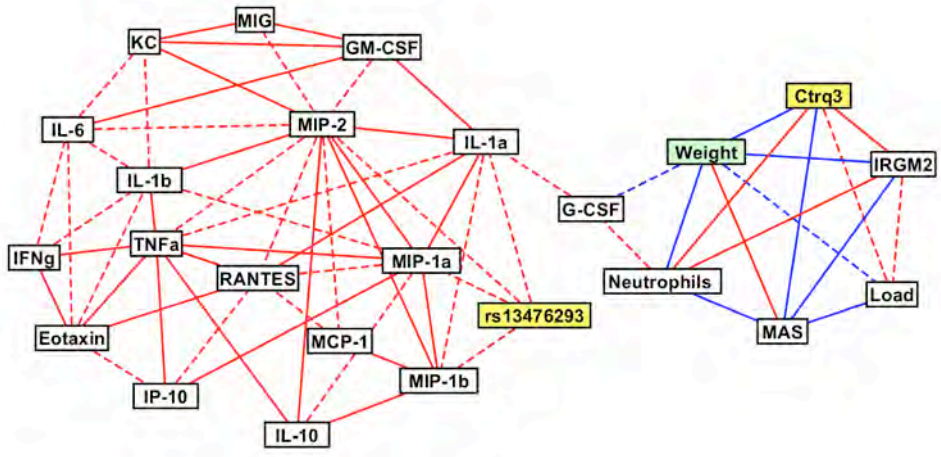


Figure 1. The correlation network of immune parameters during Chlamydia infection in BXD mice. Correlation network linking BXD genotypes (*Ctrq3* and rs13476293), *C. psittaci* load, inflammatory responses, cytokine profiles, IRGM2 protein expression pattern, and weight change after *C. psittaci* infection in BXD strains. Positive (red) and negative (blue) correlations between variables with magnitudes of Pearson's correlation coefficient greater than 0.6 (dashed lines) and 0.7 (solid lines) are shown.

The first step in our analysis was to

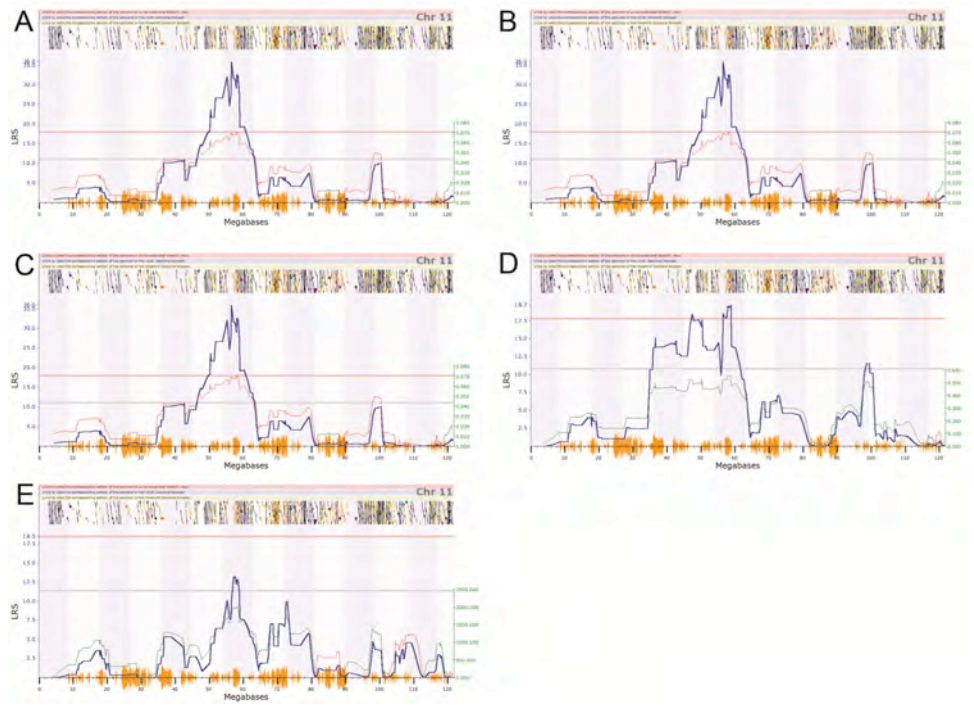


Figure 2. QTL mapping results for (A) day 6 weight, (B) neutrophils, (C) macrophage activation status, (D) *C. psittaci* load, and (E) G-CSF on chromosome 11. Significant QTLs ($p < 0.05$) are indicated by the solid red line.

determine which intermediate phenotypes to include in the model. Data for 32 cytokines, pathogen load, levels of macrophages and neutrophils, and macrophage activation status (MAS) were collected for 40 BXD strains infected with *Chlamydia psittaci*. We used two tools, correlation network analysis and QTL mapping, in the GeneNetwork web server (www.genenetwork.org) to determine which of these variables were associated with disease status. We found that many cytokines were highly correlated, but had little association with diseases status (**Figure 1**). Disease status (weight) was associated the level of the G-CSF cytokine, the level of neutrophils, pathogen load, and MAS. As expected, disease status was controlled by a QTL at *Ctrq3*. Neutrophils, load, and MAS also had significant QTLs at this location, but G-CSF did not (**Figure 2**). To further investigate the influence of the genetic polymorphisms at *Ctrq3*, we analyzed the expression pattern of the IRGM2 protein, an innate immune gene in the family of immunity-related GTPases located at *Ctrq3*, and found that it had two distinct isoforms that were highly correlated with *Ctrq3* genotype. We also investigated the protein expression level of IRGB10, but found that it did not correlate with weight. Of the 40 BXD strains studied, there was no discrepancy between the *Ctrq3* genotype and IRGM2 expression pattern. We, therefore, selected 5 variables to include in the Bayesian network model: genotype at *Ctrq3*/IRGM2 protein expression pattern, neutrophil level, pathogen load, MAS, and weight ratio.

Structural learning of the Bayesian network was performed using the R package *deal* (http://cran.r-project.org/web/packages/deal/index.html). The network was constructed from data for C57BL/6J and 40 BXD strains using one discrete node, representing the *Ctrq3* genotype and IRGM2 protein expression pattern, and four continuous nodes (neutrophils, *C. psittaci* load, macrophage activation status, and the ratio of the weight of the mice 6 days after infection to the weight before infection), which were modeled with conditional Gaussian distributions. The Bayesian network score (2), which is basically a version of the BDe scoring metric (5) extended to include conditional Gaussian distributions, was calculated in *deal* for

each possible network structure with the following restrictions: the genotype node could not be the child of any other nodes and the weight node could not be the parent of any other node. The Bayesian score metric inherently handles the problem of over fitting data to complex models (9). However, selecting a single best network model and ignoring all other models may still lead to over-fitting the data. Model averaging can be used to reduce this risk (3). An indicator function f is defined as: if a network G learned from data D has the feature (here a

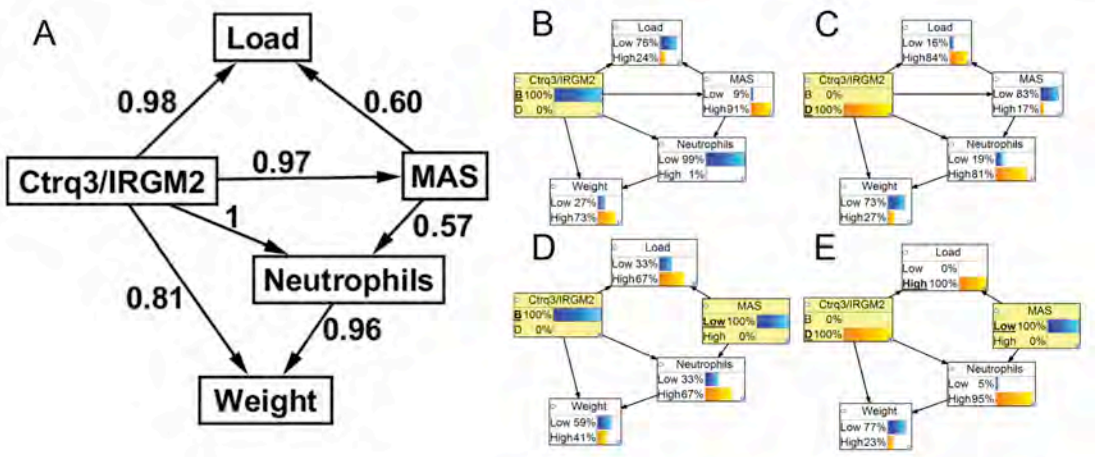


Figure 3. Bayesian network (BN) modeling of *C. psittaci* infection. (A) Structure of the BN. The number next to each directed arc of the BN indicates the confidence in the arc after model averaging, with higher numbers indicating higher confidence. (B and C) Predictions of the BN as a function of genotype at *Ctrq3*. BXD strains with the susceptible (D) genotype at the IRG locus tend to have lower MAS and weights and higher levels of neutrophils and pathogen load. (D and E) Predicted effect of interventional depletion of macrophages on the predictions of the BN. External depletion of macrophages is predicted to cause a decrease in weight and increases in neutrophils and pathogen load. For panels B to E, a threshold was determined by averaging the mean value of strains with a B genotype and the mean value of strains with a D genotype for the original data sample. Then, the probability that the predicted value for each variable was greater than (High) or less than (Low) this threshold was calculated from the conditional Gaussian distributions learned from the BN for the original data and after macrophage depletion. Nodes with a yellow background have been assigned the value to represent data for only a given genotype and intervention on the MAS.

feature is a directed edge representing a regulatory relationship), $f(G)=1$, otherwise, $f(G)=0$. The posterior probability of a feature is $P(f(G)\mid D)=\sum_{G} f(G)P(G\mid D)$. This probability reflects our confidence in the feature f. Leave-one-out cross validation was also used to test the performance of the hybrid Bayesian network. For each test strain, parameter learning of the remaining 40 strains and inference was performed with the Bayes Net Toolbox with a maximum likelihood approach. To evaluate the quality of the continuous predictions, we used the Q^2 parameter (4), which is given by:

$$Q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

where y_i is the value of the ith sample, $\hat{y}_i$ is the predicted value of the ith sample, and $\bar{y}$ is the sample mean. The values of Q^2 for MAS, neutrophil level, pathogen load, and weight were 0.51, 0.59, 0.45, and 0.68, respectively. Additionally, we discretized the original data and the

predictions from the leave-one-out-cross validation for each strain and used this discretized data to test the accuracy of the predictions. A threshold for each of the continuous variables was determined by averaging the mean value of the original data for all strains with the B genotype and the mean value for all strains with the D genotype. Then, the continuous variables for the original data and the predictions were classified as being either High or Low through comparison with these threshold values for each strain. The accuracy was then determined by dividing the number of predictions that matched the original data by the total number of strains. For MAS, neutrophil level, pathogen load, and weight, the accuracy was accuracy 85%, 93%, 80%, and 88%, respectively.

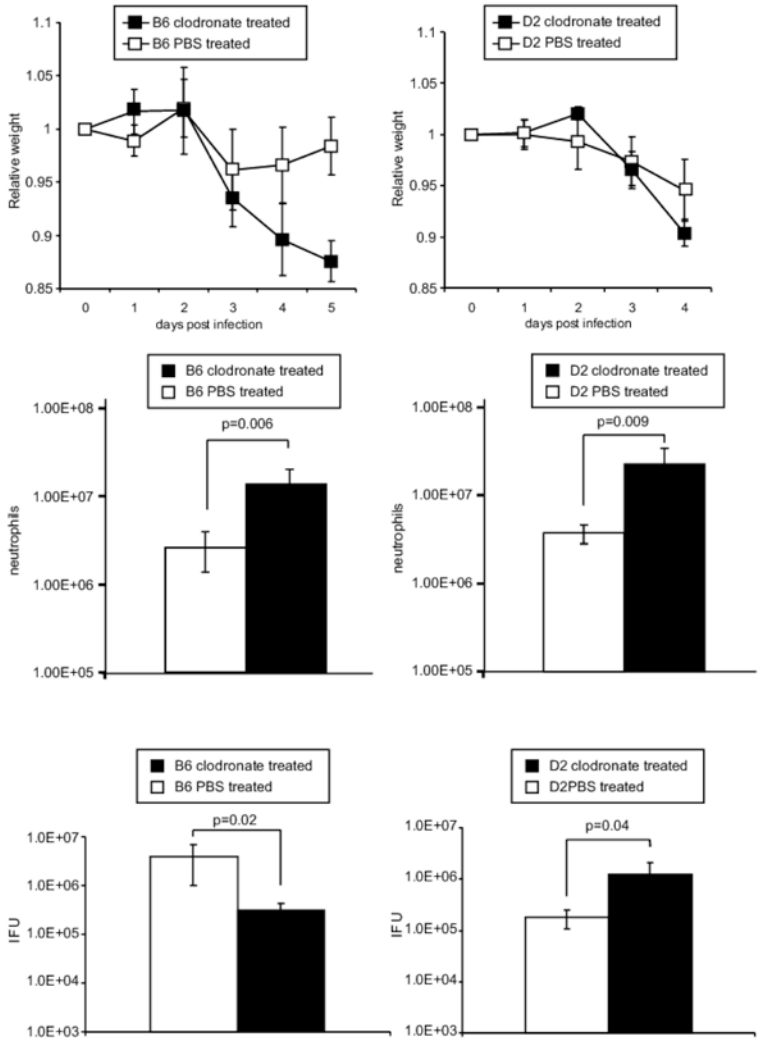


Figure 4. Impact of macrophage depletion on course of *C. psittaci* infection in B6 and D2 strains. B6 and D2 mice received either liposome clodronate or liposome PBS iv (day –1 post infection) and i.p. (day –1, 1, and 3 post infection) and infected with *C. psittaci* 10^4 IFU intraperitoneally (N=5/group). Mice were monitored daily for weight (Top) and appearance. Clodronate treated D2 strain became moribund on day 4 and B6 on day 5 post infection and were euthanized. (Middle) The number of neutrophils in the peritoneal lavage increased in both B6 and D2 strains treated with clodronate. (Bottom) Pathogen load in the peritoneal cavity was quantified. There was a statistically significant difference in final weight, number of neutrophils, and *C. psittaci* load for both B6 and D2 strains. Data is representative of two experiments.

The structure of the Bayesian network model is presented in **Figure 3**. We further investigated two of the relationships implied by the network. First, MAS is central to the model, implying that MAS directly influences neutrophils and pathogen load, and also indirectly impacts disease status. We can use the model to predict the effects of external interventions, such as depleting the level of macrophages and MAS. The intervention sets of the value of the MAS node and relieves it from the influence of its parent node. Therefore, we predict the effects of macrophage depletion by removing the link *Ctrq3*/IRGM2 → MAS

and setting MAS to the minimum value observed in the data used for parameter learning [action do (MAS = MIN), where MIN is the minimum observed MAS value]. The probabilistic inference of the effects of this influence was executed using the Bayes Net Toolbox. The model predicts that depletion of macrophages increases the levels of neutrophils and pathogen load and decreases weight (**Figure 3** panels b-e). The magnitude of these changes is expected to be more significant in BXD strains with the resistant (D) genotype.

We tested these predictions by performing chemical depletion of macrophages with clodronate before infecting B6 and D2 strains with *C. psittaci*. The results validated many of the model's predictions (**Figure 4**). In the D2 strain, depletion of macrophages increased neutrophil influx; and *C. psittaci* load and also induced a more rapid decline in weight. As predicted, in the resistant B6 strain, depletion of macrophages increased neutrophils and exacerbated the weight loss. These mice were moribund 5 days post-infection. In contrast to prediction, the B6 strain had decreased pathogen load after depleting macrophages. Pathogen load in the liver was similar in B6 irrespective of whether macrophages were depleted or not (PBS control: 4.26 x 10^6 IFU/gram, Clodronate treated; 3.81 x 10^6 IFU/gram, p=0.06).

Several reports have documented that IRGs reduce pathogen burden *in vitro* and *in vivo*, which is expected to influence disease severity (1, 6, 8). To investigate the possibility that the status of the *Irgm2* genotype switches disease modality, we performed the Bayesian analysis for strains with the B genotype at the *Ctrq3* locus separately from strains with the D genotype. As each data set only contained data from strains with one genotype, the genotype node was removed from the network, and the structure of the network, using model averaging of an exhaustive search of possible structures with deal, was learned for both the B and D data. For strains with the D genotype, a directed edge from pathogen load to weight ratio had a posterior probability of 0.64, while the same edge for strains with the B genotype had a posterior probability of only 0.16, indicating that weight change was dependent on *C. psittaci* load only for strains with the susceptible D genotype. The influence of load on disease outcome was much stronger for strains with the susceptible D genotype than strains with the resistant B genotype. Because of the genotype-specific switching between pathogen load and disease in the model, we expanded this analysis to 197 BXD mice and correlated the disease severity (weight) with pathogen load in individual mice according to the genotype at the IRG locus (**Figure 5**). Overall, mice with the B genotype had lower pathogen load compared to mice with a D

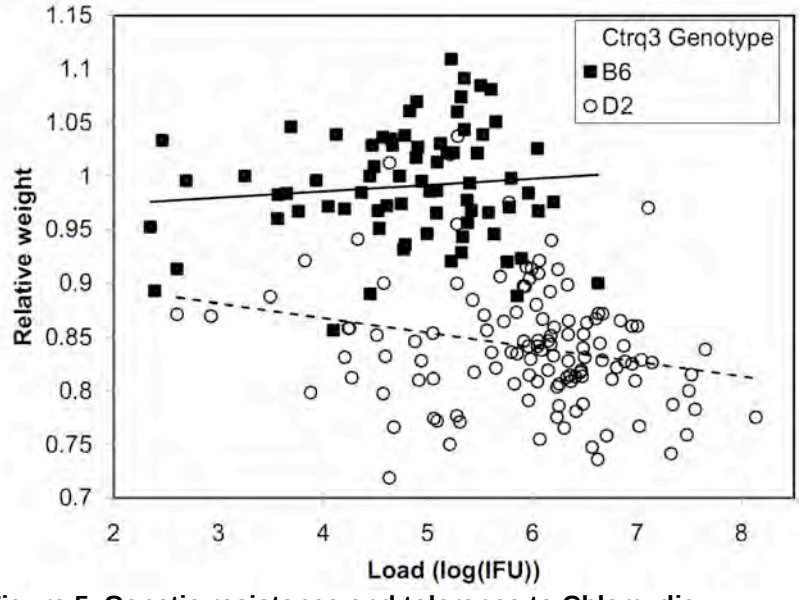


Figure 5. **Genetic resistance and tolerance to Chlamydia infection in mice.** Plot of weight change as a function of IFU for 197 BXD mice infected with *C. psittaci*. Mice with the susceptible D genotype at the *Ctrq3* (open symbols) lose weight as pathogen load increases, while mice with the resistant B genotype at the *Ctrq3* marker (filled circles) do not. The slopes of the linear regression lines for the B (solid line) and D (dashed line) data are significantly different (p = 0.02).

genotype, although a considerable overlap existed. However, the mice with the B genotype were tolerant of increases in pathogen burden whereas mice with D genotype lost more weight with increases in pathogen burden as demonstrated by the differences in the slope of the load to weight linear regression lines (p = 0.02).

Additionally, we have begun to analyze data for BXD strains infected with *Acinetobacter baumanii*, antibiotic resistant bacteria that have caused significant morbidity and mortality among military personnel with combat-associated wounds and in intensive care units. The levels of 30 cytokines in the bronchoalveolar lavage fluid (BALF) were measured in 17 BXD strains, respectively. Similar to the results for cytokines after *C. psittaci* infection, there is a large amount of correlation between the levels of different cytokines **(Figure 6**). We have also identified three genomic locations [Chr 6 near 65 Mb (mCV22576656), Chr 14 near 43 Mb (rs13482156), and Chr 19 near 51 Mb (rs3716572)] that each impact the level of several cytokines. We performed principal components analysis of the cytokines with an absolute correlation greater than 0.7 for the three loci and mapped the first principal component of each of the groups of cytokines (**Figure 7**). The first principal component of each of the sets of cytokines has a significant or highly suggestive QTL at the marker location. Several genes at these three loci, including a cluster of immunoglobulin kappa (IGK) genes on Chr 6, *Mbl1* and *Bmp4* on Chr 14, and *Nfkb2* on Chr 19, have been previously associated with immune response. These locations are promising starting points for developing network models that explain variation in response to infection with *Acinetobacter baumannii*.

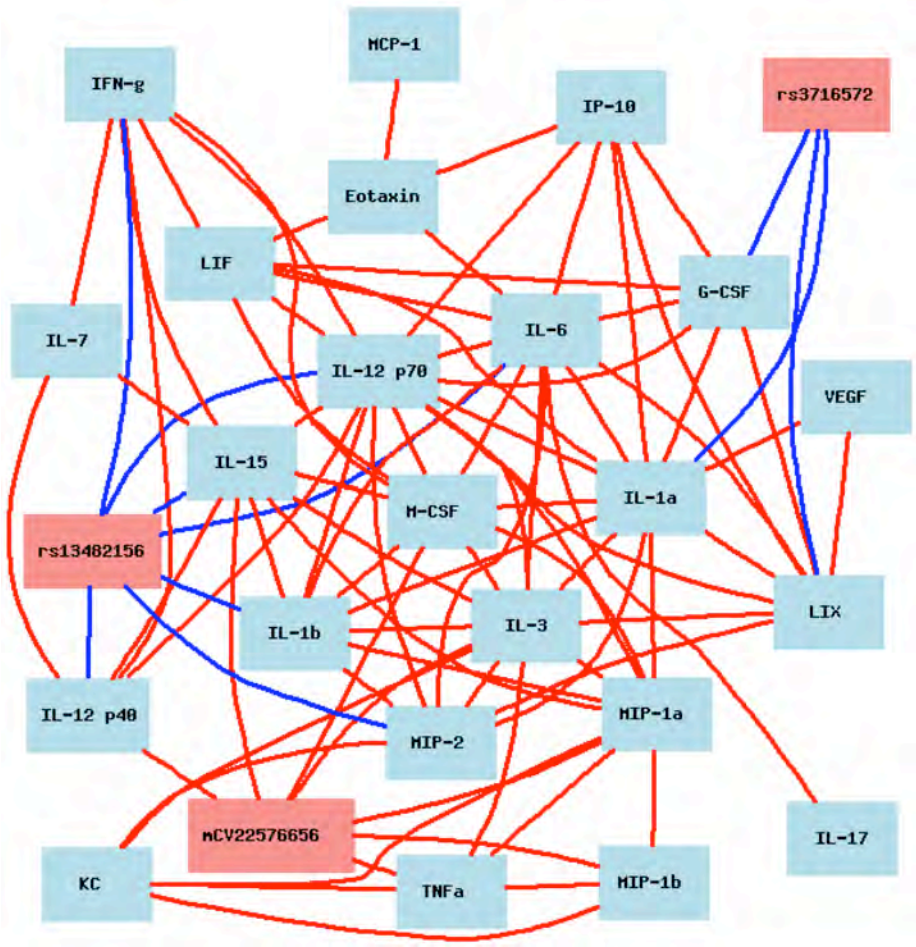


Figure 6. Correlation network for cytokine levels in the BALF of BXD strains infected with *Acinetobacter baumannii*. Three genotype markers (red boxes) on chromosomes 6, 14, and 19 are highly correlated with several cytokines. Pearson's correlation coefficient greater than 0.7 (solid lines) are shown.

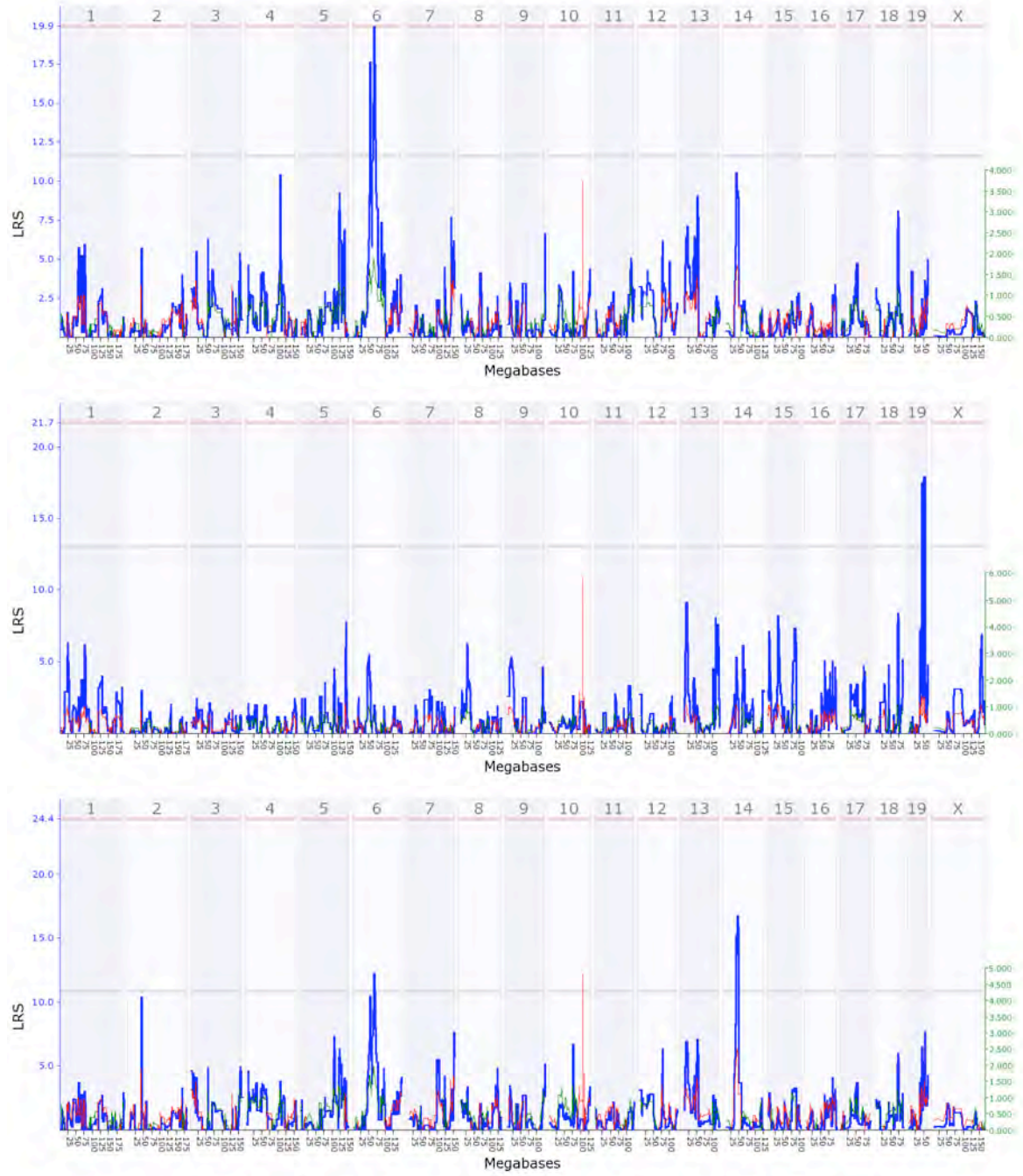


Figure 7. QTL mapping of first principal component for cytokines correlated (Pearson correlation coefficient > 0.7) with mCV22576656 (top), rs13482156 (middle), and rs3716572 (bottom).

The cytokine levels in the serum of 13 BXD strains infected with *Acinetobacter baumannii* were also collected. Again, the cytokines were highly correlated (Figure 8). However, in general, the correlation coefficients between the cytokine levels were lower in the serum than in the BALF, a

result that may have been due to the lower number of strains in the serum data set. The Chr 6 (mCV22576656) and Chr 14 (rs13482156) loci that influence the levels of cytokines in the BALF are also correlated with cytokine levels in the serum.

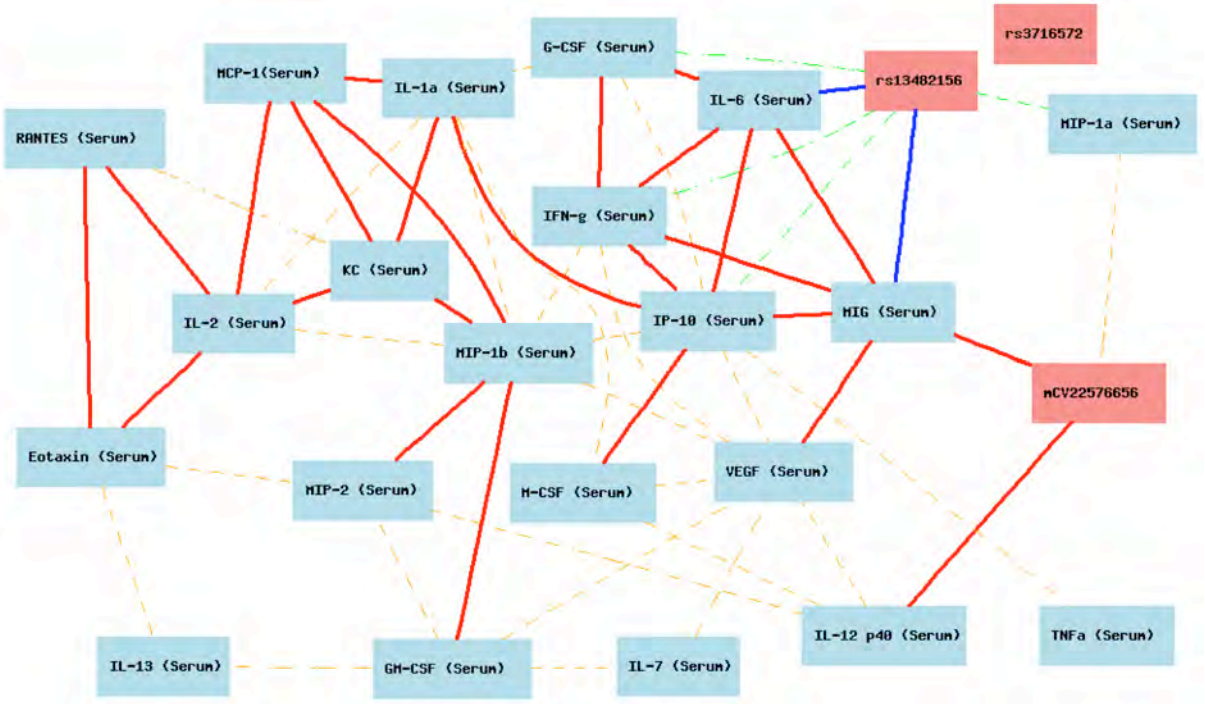


Figure 8. Correlation network for cytokine levels in the serum of BXD strains infected with *Acinetobacter baumannii*. Two of three genotype markers (red boxes) that are highly correlated with cytokine levels in the BALF are also correlated with cytokine levels in the serum are shown. Absolute value of Pearson's correlation coefficient greater than 0.7 (solid lines) and 0.6 (dashed lines) are shown.

6.3 Key Research Accomplishments

- Developed integrated method for Bayesian network structure learning, parameter learning, cross-validation, and prediction of effects of external interventions.
- Created Bayesian network model connecting *Ctrq3* genotype with disease status in BXD mice after infection with *Chlamydia psittaci*
- Used the model to predict the effects of an external intervention, specifically macrophage depletion, and experimentally verified many of these predictions
- Found that genetic influence on of *Chlamydia* disease in strains with the resistant D genotype was a function of both increased resistance to and tolerance of high pathogen loads

6.4 Reportable Outcomes

Publications

1. Bao L, Xia X and Cui Y. Expression QTL modules as functional components underlying higher-order phenotypes. PLoS One. 5(12): e14313 (see appendices).

Manuscript submitted for publication

Verma NK, Abhishek R and Cui Y. Improved Mountain Clustering Algorithms for Gene Expression Data Analysis. BMC Bioinformatics. (Under review)

Manuscript in preparation

Miyairi I, Ziebarth J, Laxton JD, Wang X, van Rooijen N, Williams RW, Lu L, Byrne G and Cui Y. Prediction and Validation of Genetic Tolerance to *Chlamydia psittaci* Infection. (In preparation)

Oral presentation

Ziebarth, J. D.; Cui, Y.; Miyairi, I. "Using hybrid Bayesian network models to link genotype with phenotype in infectious diseases". 10th Annual UT-ORNL-KBRIN Bioinformatics Summit 2011, April 3, Memphis TN.

6.5 Conclusion

We have implemented a method for performing Bayesian network analysis of pathogen infection and applied it to a model system of BXD mice infected with *Chlamydia psittaci*. We were able to confirm several of the predictions of the model with subsequent experiments. This work is significant not only because it improves understanding of the genetic basis of variation in response to *C. psittaci*, but also because it provides a workflow that can be applied to understand responses to other infectious diseases.

6.6 References

1. Bernstein-Hanley, I., J. Coers, Z. R. Balsara, G. A. Taylor, M. N. Starnbach, and W. F. Dietrich. 2006. The p47 GTPases Igtp and Irgb10 map to the Chlamydia trachomatis susceptibility locus Ctrq-3 and mediate cellular resistance in mice. Proc Natl Acad Sci U S A 103:14092-7.
2. Bottcher, S. 2001. Presented at the Eighth International Workshopin Artificial Intelligence and Statistics.
3. Hartemink, A. J., D. K. Gifford, T. S. Jaakkola, and R. A. Young. 2001. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. Pac Symp Biocomput.
4. Hawkins, D. M., S. C. Basak, and D. Mills. 2003. Assessing model fit by cross-validation. J Chem Inf Comput Sci 43:579-86.
5. Heckerman, D., D. Geiger, and D. Chickerling. 1995. Learning Baysian networks: The combination of knowledge and statistical data., p. 197-243, Machine Learning, vol. 20. Microsoft Research.
6. MacMicking, J. D., G. A. Taylor, and J. D. McKinney. 2003. Immune control of tuberculosis by IFN-gamma-inducible LRG-47. Science 302:654-9.
7. Miyairi, I., V. R. Tatireddigari, O. S. Mahdi, L. A. Rose, R. J. Belland, L. Lu, R. W. Williams, and G. I. Byrne. 2007. The p47 GTPases Iigp2 and Irgb10 regulate innate immunity and inflammation to murine Chlamydia psittaci infection. J Immunol 179:1814-24.
8. Nelson, D. E., D. P. Virok, H. Wood, C. Roshick, R. M. Johnson, W. M. Whitmire, D. D. Crane, O. Steele-Mortimer, L. Kari, G. McClarty, and H. D. Caldwell. 2005. Chlamydial IFN-gamma immune evasion is linked to host infection tropism. Proc Natl Acad Sci U S A 102:10658-63.
9. Pe'er, D. 2005. Bayesian network analysis of signaling networks: a primer. Sci STKE 2005:pl4.

Plasmid 64 (2010) 156–161

Contents lists available at ScienceDirect

Plasmid

journal homepage: www.elsevier.com/locate/yplas

Construction of a bioluminescence reporter plasmid for *Francisella tularensis*

Xiaowen R. Bina, Mark A. Miller, James E. Bina *

The University of Tennessee Health Science Center, Department of Molecular Sciences, 858 Madison Avenue, Memphis, TN 38163, USA

ARTICLE INFO

Article history:
Received 21 March 2010
Accepted 4 July 2010
Available online 8 July 2010

Communicated by J. Casadesus

Keywords:
Francisella tularensis
Plasmid
Luminescence

ABSTRACT

A *Francisella tularensis* shuttle vector that constitutively expresses the *Photorhabdus luminescens* lux operon in type A and type B strains of *F. tularensis* was constructed. The bioluminescence reporter plasmid was introduced into the live vaccine strain of *F. tularensis* and used to follow *F. tularensis* growth in a murine intranasal challenge model in real-time by bioluminescence imaging. The results show that the new bioluminescence reporter plasmid represents a useful tool for tularemia research that is suitable for following *F. tularensis* growth in both in vitro and in vivo model systems.

1. Introduction

Francisella tularensis is a gram negative facultative intracellular bacterium that causes the zoonotic disease tularemia. *F. tularensis* infection of humans can occur by a number of routes, including the handling of infected animals, arthropod bites (Evans, 1985; Francis, 1937; Tarnvik, 1989), ingestion (Anda et al., 2001; Greco et al., 1987; Karpoff and Antononoff, 1936), and by inhalation (Syrjala et al., 1985; Teutsch et al., 1979). *F. tularensis* is highly infectious and as few as 10 bacteria can cause disease (Cross, 2000). The high infectivity and ease of dissemination of *F. tularensis* by aerosols has raised concerns about the potential use of *F. tularensis* as a biological weapon (Sjostedt, 2007) and provided the rationale for the development of new tularemia therapeutics.

A major focus of *F. tularensis* research is to decipher the molecular mechanisms that contribute to *F. tularensis* pathogenesis. The strategy for identification of virulence associated genes has largely focused on generating mutations in putative virulence genes and assessing the resultant strains for growth attenuation in a murine tularemia model. The traditional method for assessing *F. tularensis* growth and dissemination in vivo requires challenging large numbers of animals with the test and control *F. tularensis* strains. Thereafter, several animals are sacrificed from each group and dissected at each time point over a time course. *F. tularensis* titers are then determined in each mouse by plating serial dilutions of organ homogenates onto agar plates. This method requires large numbers of experimental animals and is laborious. In addition, the requirement for repeated animal sacrifice, dissection and tissue handling increases the potential for occupational exposure of researchers to *F. tularensis*, a category A select agent.

Technological advances in small animal imaging have made it possible to monitor in real-time the growth and dissemination of fluorescent or bioluminescent-labeled bacteria in individual animals over the entire course of infection, offering a powerful alternative to traditional methodologies. Bioluminescence has proven to be particularly useful for this application. Bioluminescence reporters have several advantages over fluorescence reporters for in vivo imaging studies. Luminescence reporters are more sensitive and have lower background levels, and they do not share the auto-fluorescence or signal quenching issues that limit the utility of fluorescent reporters for in vitro

* Corresponding author. Fax: +1 901 448 7360.
E-mail address: JBina@uthsc.edu (J.E. Bina).

0147-619X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.plasmid.2010.07.001

and in vivo imaging applications. In addition, there are numerous methods available for the detection of bioluminescence (e.g. CCD camera, plate reader, film exposure, scintillation counter). Bioluminescence reporters have a number of applications in bacterial pathogenesis including the quantification of gene expression, virulence analysis, and the evaluation of therapeutic agents. Bioluminescence tagging vectors that express the *Photorhabdus luminescens* lux operon have been used to follow in real-time the growth and dissemination of a number of pathogens in animal models. However, to date bioluminescence reporters have not yet been published for use in *F. tularensis*.

In this work we describe the construction of a new *F. tularensis* reporter plasmid that constitutively expresses the *P. luminescens* lux operon. We show that the presence of this plasmid in type A and type B *F. tularensis* results in bioluminescence production which could be used to follow *F. tularensis* growth and dissemination in vitro and in vivo.

2. Materials and methods

2.1. Bacterial strains and growth conditions

F. tularensis strains LVS (live vaccine strain) and Schu S4 were obtained from the Centers for Disease Control and Prevention (CDC, Atlanta, GA). All work involving Schu S4 was performed in a CDC-approved BSL3 facility at The University of Tennessee Health Sciences Center in accordance with approved BSL3 protocols. *F. tularensis* strains were cultured in modified Mueller Hinton broth (MH broth supplemented with 10 g/L tryptone, 0.1% glucose, 0.025% ferrous pyrophosphate, 0.1% L-cysteine, and 2.5% calf serum) or on BHI-chocolate agar (BHI agar supplemented with 1% hemoglobin and 1% IsoVitalex). *Eschericia coli* strain EC100λpir (Epicentre, Madison, WI) was used as a host for the cloning experiments and was grown in Luria–Bertani (LB) broth or on LB agar at 37 °C. Antibiotics were used at the following concentrations when necessary: kanamycin (km) at 50 μg/mL for *E. coli* and 10 μg/mL for *F. tularensis*; cefprozil was used at 300 μg/mL for *F. tularensis*; carbenicillin was used at 100 μg/mL for *E. coli*.

2.2. Recombinant DNA methods

Recombinant DNA methods were performed according to standard protocols. Restriction enzymes were purchased from New England Biolabs (Beverly, MA). PCR amplification was performed using Biolase DNA polymerase (Bioline, Taunton, MA) or Pfu DNA polymerase (Stratagene, Cedar Creek, TX). *F. tularensis* was transformed by electroporation as previously described (Bina et al., 2006) except that outgrowth following electroporation was limited to one hour (for Schu S4) or two hours (for LVS) before plating onto selective media.

2.3. Construction of pXB173-lux

The *Francisella–E. coli* shuttle vector pXB167 (Bina et al., 2006) was used as a starting template for construction of pXB173-lux. The initial step in construction was to replace the ColE1 origin of replication with a cassette encoding the R6K origin of replication and the conjugal origin (*oriT*) from pBSL238 (Alexeyev and Shokolenko, 1995). This was accomplished by digestion of pXB167 with AclI and PacI restriction endonucleases to remove the ColE1 origin of replication. The resulting 4090 bp fragment was made blunt-ended by treatment with the Klenow fragment of DNA polymerase before being ligated to the 784 bp *ori*R6 k and *oriT* PCR amplicon that was obtained from pBSL238 by PCR using the *oriF* (5′-CGATCTACTATGCCATGTCAGCCGTTAAGTGTTCC-3′) and *oriR* (5′-GGGATATCGGGGATCAATTCCGTGATAGGTGG-3′) primers to produce pXB168 (Fig. 1).

We then replaced the *gfp* gene in pXB168 with the *aph*3′ gene that encoded kanamycin resistance. This was accomplished by digestion of pXB168 with BamHI and ClaI restriction enzymes to remove the *gfp* allele. The resulting 4119 bp fragment was rendered blunt-ended by treatment with Klenow fragment before being ligated to the 901 bp kanamycin resistance gene (*aph*3′) which was obtained from TN:EZ by PCR using the (Epicentre, Madison, WI) using the KanF (5′-AAGGCGCGCCACGCGTAGGAGTTTGT TATGAGCCATATTCAACGGGAA-3′) and KanR (5'-GCACGCG TCAAGTCAGCGTAATGCTCTGCCAG-3') primers to generate pXB169. pXB173-lux was then generated by digestion of

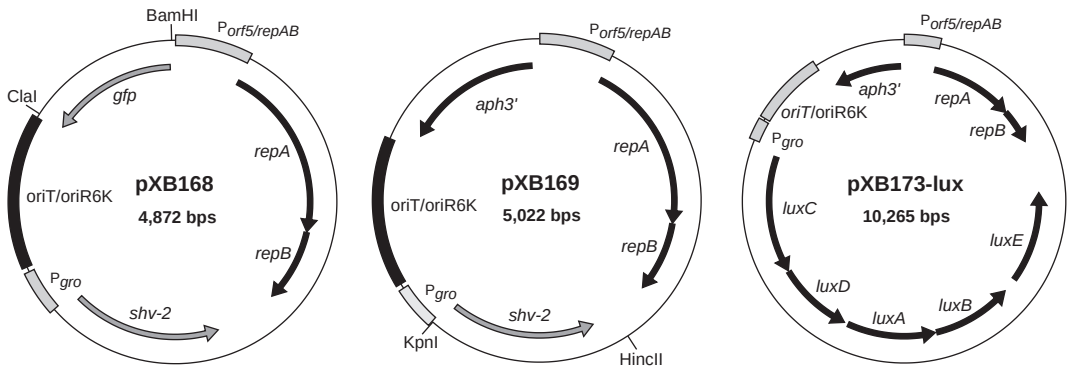


Fig. 1. Construction of pXB173-lux. Plasmid pXB168 is derived from pXB167 (Bina et al., 2006). Only relevant restriction sites are shown. The details for construction are given in Section 2.3.

pXB169 with KpnI and HinCII restriction enzymes to remove the *shv-2* allele. The resulting 3987 bp fragment was then ligated with the 5.9 kb lux operon that was derived from digestion of pXB128-lux (Bina laboratory collection) with KpnI and XmaI restriction enzymes. The results of this ligation placed the lux operon downstream and in the same orientation as the constitutively expressed Francisella *gro* promoter (indicated as P_{gro} in Fig. 1). The DNA sequence of pXB173-lux was confirmed by DNA sequencing at the Molecular Resource Center of The University of Tennessee Health Science Center (Memphis, TN). The DNA sequence of pXB173-lux has been deposited in Genbank with the accession number HM017829.

2.4. Bioluminescence detection and animal challenge studies

The limit of bioluminescence detection of LVS-pXB173-lux was determined by making serial LVS-pXB173 culture dilutions in white 96-well microtiter plates. Bioluminescence production was then quantified by use of an IVIS Spectrum imaging system (Caliper Life Sciences, Hopkinton, MA) according to the manufacturer's directions.

The utility of pXB173-lux in *F. tularensis* was assessed in a murine intranasal challenge model as previously described (Lavine et al., 2007). Briefly, 12 week-old female BALB/c mice were challenged intranasally with $\sim 5 \times 10^5$ CFU of *F. tularensis* LVS-pXB173-lux in a total volume of 50 μl that was administered as 25 μl per naris. Bioluminescence was then used as a reporter to following bacterial dissemination starting at 3 h post-challenge and then at 24-h intervals until the conclusion of the experiment. Bioluminescence production in the mice was quantified by use of the IVIS Spectrum imaging system according to the manufacturer's directions.

Bioluminescence production in Type A *F. tularensis* strain Schu S4 was assessed by use of the IVIS Spectrum to image an agar plate that had been inoculated with both the LVS and Schu S4 strains of *F. tularensis* bearing the pXB173-lux vector. Due to BSL3 restrictions we are currently unable to image mice infected with *F. tularensis* Schu S4 on the IVIS imaging system.

2.5. Plasmid stability determination

The stability of pXB173-lux in *F. tularensis* Schu S4 was assessed in vitro as follows. A fresh overnight culture of *F. tularensis* Schu S4-pXB173 was successively cultured in MH broth without km for four days. The presence of the plasmid was then determined by plating serial dilutions of the culture on days 1 and 4 onto BHI-chocolate agar plates with and without km. The ratio of km-resistant to km-sensitive bacteria was then calculated to determine pXB173-lux stability in the absence of antibiotic selection. In vivo stability of pXB173-lux was determined by intranasal infection of mice with $\sim 10^3$ cfu of *F. tularensis* Schu S4-pXB173. The spleen was then collected from one mouse on days 1, 5 and 6 and homogenized in 1 ml of PBS before 0.25 ml of 5× disruption buffer (2.5% saponin, 15% BSA, in PBS) was added with light vortexing. Serial dilutions of the spleen homogenates were then plated onto BHI-chocolate agar plates with and without km. The ratio of km-resistant to km-sensitive bacteria was then calculated as an indicator of in vivo plasmid stability.

3. Results and discussion

3.1. Construction of pXB169

We previously described the construction of three shuttle vectors (pXB136, pXB160 and pXB167) (Bina et al., 2006) that were derived from pFNLTP6::gfp (Maier et al., 2004). In electroporation experiments we observed that pXB136 could be efficiently transformed into Schu S4 with selection for cefprozil resistance, however, we were not able to recover Schu S4 transformants when selecting for kanamycin resistance (data not shown). Since the kanamycin resistance locus in pXB136 is derived from the pFNLTP shuttle vectors, this observation is consistent with previous findings that pFNLTP-based vectors transformed poorly into Schu S4 (LoVullo et al., 2006) and suggested that the poor transformation and plasmid instability observed in pXB136 was likely due to inefficient expression of the kanamycin resistance allele in type A *F. tularensis* strains. In silico analysis of pXB136 suggested that the *repAB* locus in pXB136 (and in the pFNLTP vectors) contained a divergently transcribed promoter, denoted as P_{orf5} in Fig. 1, located 808 bp upstream of the kanamycin resistance gene (i.e. *aph*3′). Downstream of the P_{orf5} is *orf*5′ which encodes a truncated gene that was hypothesized to form part of a two component toxin–antitoxin system that was present in the parent plasmid pFNL10 (Pavlov et al., 1996). Downstream of *orf5*' was the f1 origin of replication and the *aph*3′ gene which originated from pCR2.1-TOPO (Maier et al., 2004). As P_{orf5} was derived from pFNL10, we hypothesized that it likely encoded an active *F. tularensis* promoter and contributed to *aph*3′ expression in pXB136 and that the intervening 808 bp sequence inhibited *aph*3′ expression in Schu S4. To test this hypothesis we deleted the 808 bp intervening region. The resulting plasmid retained kanamycin resistance in *E. coli* and LVS and gained the ability to be retained by Schu S4. The resulting plasmid transformed into Schu S4 at an efficiency that was equivalent to what was observed with LVS ($\sim 10^5$ cfu/μg DNA). Collectively these results suggested that the *repAB* promoter region contained a divergently transcribed promoter that was constitutively expressed in both *E. coli* and *F. tularensis*. It is unclear why pXB136 and the pFNLTP plasmids display different stabilities in LVS and Schu S4 with selection for kanamycin resistance.

3.2. Construction of pXB173-lux

Having established that the Schu S4 stability problems associated with our previous vectors was likely due to expression of the kanamycin resistance allele and not some inherent problem with the plasmid construct, we set out to design a new shuttle vector that could be used as a bioluminescence reporter in *F. tularensis*. To construct the bioluminescence reporter plasmid we first replaced the high copy number origin of replication that was present in pXB167 with a low copy number origin of replication and a conjugal origin of transfer. This was accomplished by

replacement of the pXB167 ColE1 origin of replication with a cassette that encoded the R6K origin of replication and the RP4 origin of transfer to generate pXB168. This effectively reduced the plasmid copy number in *E. coli* and introduced an origin of transfer to facilitate conjugal transfer of the plasmid into *F. tularensis*. Conjugation represents a very efficient and easy method for introduction of plasmids into *F. tularensis*. The *shv*-2 marker was then replaced with the *aph*3′ allele as kanamycin resistance is the most reliable and widespread genetic marker used in type A *F. tularensis* strains (e.g. Schu S4). We used the *orf5* promoter to drive expression of *aph*3′ so that we could use the P_{gro} promoter to drive expression of the lux reporter construct (see below). The resulting plasmid, pXB169, was transformed into Schu S4 with high efficiency ($\sim 10^5$ transformants per μg/DNA).

The bioluminescence reporter plasmid pXB173-lux (Fig. 1) was then generated from pXB169 by cloning the *P. luminescens* lux operon downstream of the *F. tularensis gro* promoter. The *P. luminescens* lux operon contains the genes that are required for production of both luciferase (*luxAB*) and luciferin (*luxCDE*) and expression of the lux operon results in concomitant light production. Since the *F. tularensis gro* promoter is constitutively expressed in *E. coli* and *F. tularensis*, the presence of pXB173-lux in *E. coli*, LVS and Schu S4 results in constitutive bioluminescence production as observed in Fig. 2. The in vitro detection limit for LVS-pXB173-lux in white 96-well microtiter plates was ~2000 cfu per well which suggests that pXB173-lux likely can be used to follow *F. tularensis* growth in cell culture studies.

3.3. Use of pXB173-lux to follow F. tularensis dissemination in mice

We documented the utility of pXB173-lux by testing whether it could be used as a reporter to follow *F. tulanensis* growth in a murine model of tularemia in real-time. We therefore challenged two BALB/c mice with $\sim 10^5$ cfu of LVS-pXB173-lux by the intranasal route (Fig. 3). In the first mouse, the majority of the LVS inoculum was detected in the stomach at 3 h post-challenge, suggesting that at least a portion of the intranasal challenge dose failed to reach the lungs and was swallowed. Twenty-four-hours later, the bioluminescence production in the stomach of mouse one had resolved and LVS was clearly visualized in the lungs of both animals and in the upper respiratory tract of mouse two. The upper airway infection intensified over the six day course of the experiment in mouse two. It is unclear whether this upper respiratory tract infection occurs in natural inhalation infections or is an artifact of the intranasal inoculation method that is widely used by the tularemia research community. The images also showed that LVS disseminated to the cervical lymph nodes of mouse one (day 3) and mouse two (day 2) and that colonization of the lymph nodes intensified throughout the study period. Beginning on day 3 post-challenge LVS was observed in the liver of both mice, and by day 4 post-challenge, the

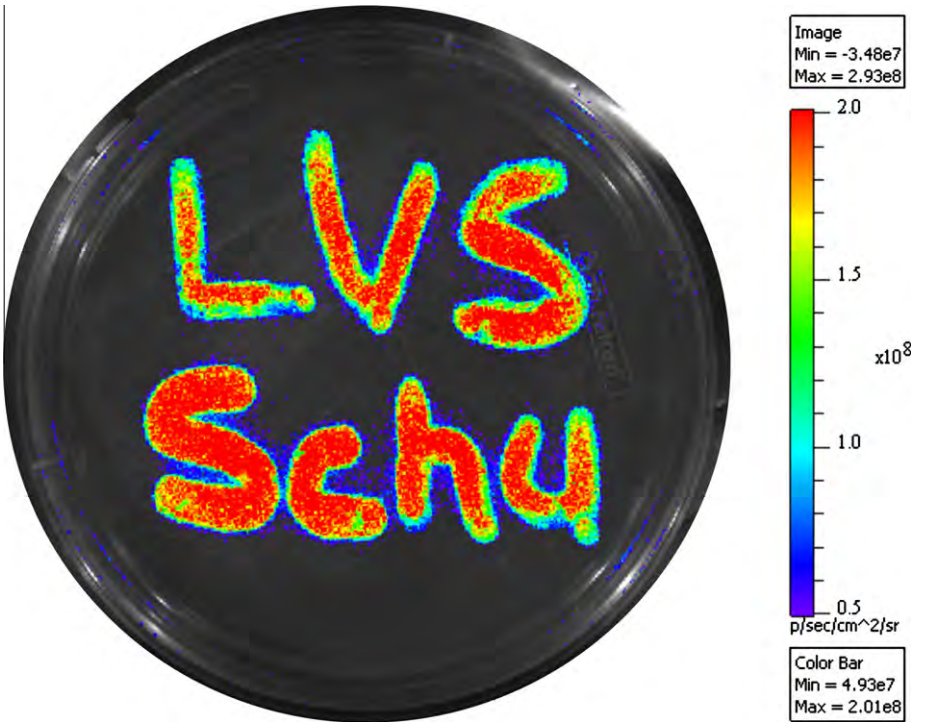


Fig. 2. Bioluminescence production by *F. tularensis*. Overnight cultures of *F. tularensis* containing pXB173-lux (LVS on the upper half of plate; Schu S4 on the lower half of plate) were inoculated onto the surface of a modified Mueller–Hinton agar plate using a Dacron-tipped swab and incubated at 37 °C for 18 h when the plate was imaged for bioluminescence production using an IVIS spectrum imaging system. Photon emission intensity is represented as a pseudocolor image that is superimposed onto the surface of the inoculated agar plate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

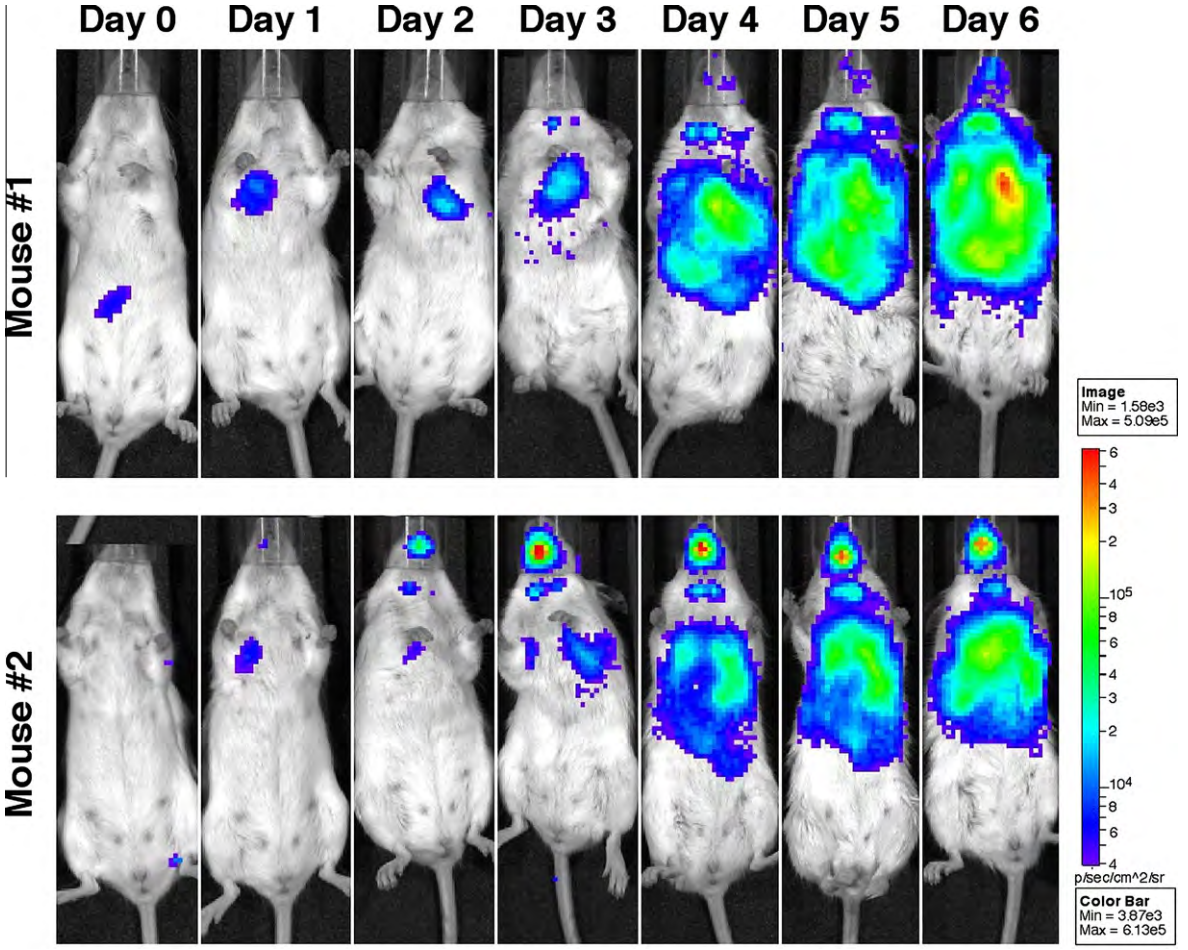


Fig. 3. Visualization of *F. tularensis* LVS-pXB173-lux in mice by bioluminescence imaging. Two twelve week-old BALB/c mice were challenged with 5×10^5 CFU *F. tularensis* LVS-pXB173-lux in a total volume of 50 μl of PBS via the intranasal route. Bioluminescence production in the mice was then visualized using an IVIS Spectrum Imaging system at 24-h intervals beginning 3 h after administration of the challenge dose. Exposure times varied based on bioluminescent signal intensities in an effort to collect between 600 and 60,000 counts, and image scaling was normalized by converting total counts to photons/s. Results shown here are representative of several experiments of similar design. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

livers of both mice were heavily colonized. On day 6, both mice exhibited extensive bacterial dissemination which correlated with other signs of severe tularemic disease (i.e. significant weight loss, ruffled fur and reduced physical activity) and the experiment was terminated. These results validate that pXB173-lux can be used as a reporter to follow *F. tularensis* dissemination in mice.

The stability of pXB173-lux in *F. tularensis* Schu S4 was assessed to validate the use of this plasmid in a Type A strain background. The plasmid was well maintained in vitro with 85% of the bacteria retaining the plasmid following growth for four successive subcultures in the absence of antibiotic selection. In vivo stability was similar to the in vitro results with 82% and 75% and of the bacteria retaining the plasmid on days 5 and 6 post-challenge, respectively. This demonstrates that pXB173-lux is stable in *F. tularensis* Schu S4, while the data presented in Fig. 2 demonstrate that the lux reporter works in *F. tularensis* Schu S4. Collectively these results strongly suggest that pXB173-lux will be useful for in vivo studies with Type A *F. tularensis* strains as we have documented with *F. tularensis* LVS.

The results presented above show that bioluminescence is a highly sensitive reporter that can be used to follow *F. tularensis* growth in mice in real-time. Bioluminescence represents a new tool for the tularemia research community that has not been previously available. In particular, the use of pXB173-lux can greatly facilitate animal and cell culture studies with virulent type A *F. tularensis* strains. As most analyses previously depended on terminal end point assays, the use of bioluminescence should greatly reduce both the labor cost and number of animals that are required for these assays while limiting the potential for occupational exposure of researchers to a potentially fatal pathogen.

Acknowledgments

This work was supported by NIH grant #U54 AI057157 from Southeastern Regional Center of Excellence for

Emerging Infections and Biodefense, by NIH grants AI074582 and AI079482 (to JEB), AI061260 (to MAM), and by DOD grant W81XHW-05-1-0227. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

Alexeyev, M.F., Shokolenko, I.N., 1995. RP4 oriT and RP4 oriT-R6K oriV DNA cassettes for construction of specialized vectors. Biotechniques 19, 22–24, 26.

Anda, P. et al., 2001. Waterborne outbreak of tularemia associated with crayfish fishing. Emerging Infectious Diseases 7, 575–582.

Bina, X.R. et al., 2006. The Bla2 beta-lactamase from the live-vaccine strain of *Francisella tularensis* encodes a functional protein that is only active against penicillin-class beta-lactam antibiotics. Archives of Microbiology 186, 219–228.

Cross, T.J.A.R.L.P., 2000. *Francisella tularensis* (tularemia). In: Mandell, G.L., Bennett, J.E., Dolin, R. (Eds.), Principles and Practice of Infectious Diseases. Churchill Livingston, Philadelphia.

Evans, M.E., 1985. *Francisella tularensis*. Infection Control 6, 381–383.

Francis, E., 1937. Sources of infection and seasonal incidence of tularemia in man. Public Health Reports 52, 103.

Greco, D. et al., 1987. A waterborne tularemia outbreak. European Journal of Epidemiology 3, 35–38.

Karpoff, S.P., Antononoff, N.I., 1936. The spread of tularemia through water as a new factor in its epidemiology. Journal of Bacteriology 32, 243.

Lavine, C.L. et al., 2007. Immunization with heat-killed *Francisella tularensis* LVS elicits protective antibody-mediated immunity. European Journal of Immunology 37, 3007–3020.

LoVullo, E.D. et al., 2006. Genetic tools for highly pathogenic *Francisella tularensis* subsp. *tularensis*. Microbiology 152, 3425–3435.

Maier, T.M. et al., 2004. Construction and characterization of a highly efficient *Francisella* shuttle plasmid. Applied Environmental Microbiology 70, 7511–7519.

Pavlov, V.M. et al., 1996. Cryptic plasmid pFNL10 from *Francisella novicida*-like F6168: the base of plasmid vectors for *Francisella tularensis*. FEMS Immunology and Medical Microbiology 13, 253–256.

Sjostedt, A., 2007. Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. Annals of the New York Academy of Sciences 1105, 1–29.

Syrjala, H. et al., 1985. Airborne transmission of tularemia in farmers. Scandinavian Journal of Infectious Diseases 17, 371–375.

Tarnvik, A., 1989. Nature of protective immunity to *Francisella tularensis*. Reviews of Infectious Diseases 11, 440–451.

Teutsch, S.M. et al., 1979. Pneumonic tularemia on Martha's Vineyard. New England Journal of Medicine 301, 826–828.

BMC Bioinformatics

ORAL PRESENTATION Open Access

High-throughput sequencing of the DBA/2J mouse genome

Xusheng Wang[1], Richa Agarwala[2], John A Capra[3], Zugen Chen[4], Deanna M Church[2], Daniel C Ciobanu[5], Zhengsheng Li[1], Lu Lu[1], Khyobeni Mozhui[1], Megan K Mulligan[1], Stanley F Nelson[4], Katherine S Pollard[3], Williams L Taylor[1], Donald B Thomason[1], Robert W Williams[1*]

From UT-ORNL-KBRIN Bioinformatics Summit 2010
Cadiz, KY, USA. 19-21 March 2010

Background

The DBA/2J mouse is not only the oldest inbred strain, but also one of the most widely used strains. DBA/2J exhibits many unique anatomical, physiological, and behavior traits. In addition, DBA/2J is one parent of the large BXD family of recombinant inbred strains [1]. The genome of the other parent of this BXD family—C57BL/6J—has been sequenced and serves as the mouse reference genome [2]. We sequenced the genome of DBA/2J using SOLiD and Illumina high throughput short read protocols to generate a comprehensive set of ~5 million sequence variants segregating in the BXD family that ultimately cause developmental, anatomical, functional and behavioral differences among these 80+ strains.

Results

We generated approximately 13.2 and 38.9× whole-genome short reads of DBA/2J females using Illumina GA2 and ABI SOLiD massively parallel DNA sequencing platforms. Comparing to the C57BL/6J reference genome sequence, we identified over 4.5 million single nucleotide polymorphisms (SNPs), including 84 nonsense and ~11,000 missense mutations, 78% of which are novel. We also detected ~568,000 insertions and deletions (indels) within single short reads and ~9,400 between mate-paired reads. Approximately 300 inversions were detected by SOLiD mate-pair reads, 46 of which span at least one exon. In addition, we identified ~22,000 copy number variants (CNVs) in the range of 1 Kb to 100 Kb (Figure 1).

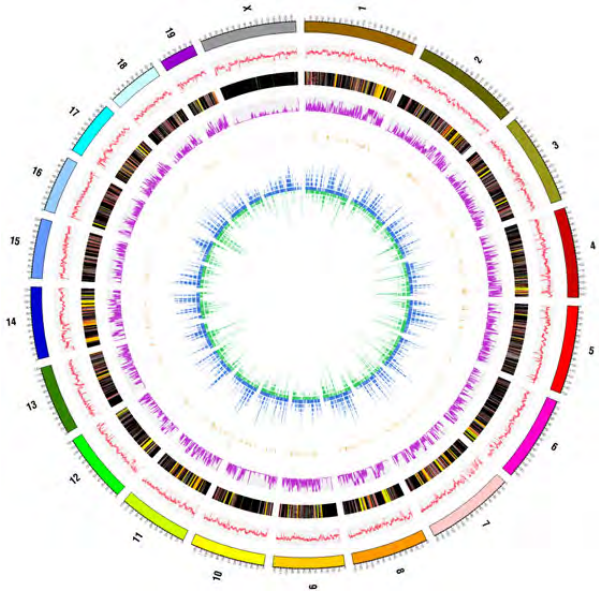


Figure 1 Concentric circles represent the sequence and structural variation across mouse chromosomes. Moving inward from the outer circle, circle 1 denotes each chromosome. Circle 2, read depth with 100kb window. Circle 3, SNP density with 100kb windows (black is lowest density and orange is highest density). Circle 4, Indels density with 100kb window. Circle 4, Inversion. Circle 5, CNVs, blue (outward) denotes loss of CNVs and green (inward) denotes gains of CNVs.

Conclusion

Our study generates the first consensus sequence for the DBA/2J and creates a compendium of sequence and structural variations that will be used by the community of researchers who study complex traits in mouse models. The sequence data provide a novel resource with which to initiate reverse genetic analysis of complex

* Correspondence: rwilliam@nb.uthsc.edu
[1]University of Tennessee Health Science Center, Memphis, TN 38163, USA

traits, particularly by exploiting strong alleles (premature stop codons, frame-shift mutations, and deletion) that differentially affect members of the BXD strain family. The DBA/2J genome is also an essential prerequisite to unbiased alignment of RNA-seq and ChIP-seq data generated using BXD strains and any other cross involving these two common parental strains.

Author details
[1]University of Tennessee Health Science Center, Memphis, TN 38163, USA. [2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. [3]Gladstone Institutes, University of California, San Francisco, CA 94158, USA. [4]University of California, Los Angeles, CA 90095, USA. [5]University of Nebraska, Lincoln, NE 68588, USA.

Published: 23 July 2010

References

1. Peirce JL, Lu L, Gu J, Silver LM, Williams RW: **A new set of BXD recombinant inbred lines from advanced intercross populations in mice.** *BMC genetics* 2004, **5**:7.
2. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420(6915)**:520-562.

doi:10.1186/1471-2105-11-S4-O7
Cite this article as: Wang *et al.*: High-throughput sequencing of the DBA/2J mouse genome. *BMC Bioinformatics* 2010 **11**(Suppl 4):O7.

OPEN ACCESS Freely available online

PLoS one

Expression QTL Modules as Functional Components Underlying Higher-Order Phenotypes

Lei Bao[1]*¤, Xuefeng Xia[3], Yan Cui[1,2]*

1 Department of Molecular Sciences, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **2** Center for Integrative and Translational Genomics, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **3** Institute of Bioinformatics, Tsinghua University, Beijing, China

Abstract

Systems genetics studies often involve the mapping of numerous regulatory relations between genetic loci and expression traits. These regulatory relations form a bipartite network consisting of genetic loci and expression phenotypes. Modular network organizations may arise from the pleiotropic and polygenic regulation of gene expression. Here we analyzed the expression QTL (eQTL) networks derived from expression genetic data of yeast and mouse liver and found 65 and 98 modules respectively. Computer simulation result showed that such modules rarely occurred in randomized networks with the same number of nodes and edges and same degree distribution. We also found significant within-module functional coherence. The analysis of genetic overlaps and the evidences from biomedical literature have linked some eQTL modules to physiological phenotypes. Functional coherence within the eQTL modules and genetic overlaps between the modules and physiological phenotypes suggests that eQTL modules may act as functional units underlying the higher-order phenotypes.

Citation: Bao L, Xia X, Cui Y (2010) Expression QTL Modules as Functional Components Underlying Higher-Order Phenotypes. PLoS ONE 5(12): e14313. doi:10.1371/journal.pone.0014313

Editor: Cathal Seoighe, National University of Ireland Galway, Ireland

Received July 22, 2010; **Accepted** November 23, 2010; **Published** December 13, 2010

Funding: This work was supported in part by NIH grants NR009270, AI081050, a Department of Defense grant W81XHW-05-01-0227 and an American Heart Association grant 0830134N. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lebao@ucsd.edu (LB); ycui2@uthsc.edu (YC)

¤ Current address: Moores Cancer Center, University of California San Diego, San Diego, California, United States of America

Introduction

Recent advances in the integration of quantitative genetics and expression genomics have provided a global view of gene expression traits and their implications in high-order phenotype variations [1,2,3,4,5,6,7,8]. The Genetical Genomics [9] approach systematically associates gene expression traits with regulatory genomic regions called expression quantitative trait loci (eQTLs) [10]. Typically, this high-throughput approach identifies a large set of regulatory relations between genetic markers and expression traits, which compose bipartite networks that consist of two types of nodes, representing expression traits and eQTLs respectively.

A module is usually defined as a subset of components in a network that interact with each other and act in concert to regulate biological processes, while maintaining relative independence from other components in the network. Studies on the architecture of biological networks, including protein-protein interaction networks, metabolic networks, and transcriptional regulatory networks [11,12,13] have revealed that modularity is a common organizational principle of these networks. In a previous work we discovered transcription modules and their associations with higher-order phenotypes [14]. Recently a Bayesian method for eQTL network partition was developed by Zhang et al. [15]. The application of their method to a yeast eQTL network identified 20 modules with one eQTL and 9 modules with two eQTLs [15].

In this work we define eQTL module as a set of highly connected nodes with at least two eQTLs in different chromosomes. We analyzed the eQTL networks constructed from a yeast dataset and a mouse liver dataset and found 65 and 98 modules respectively. We also studied the associations between the eQTL modules and higher-order phenotypes. Genes in many eQTL modules showed significant functional coherence. Fifty yeast morphologic phenotypes were mapped to genetic loci that overlapped with the eQTLs in 19 modules. We identified an eQTL module sharing genetic components with a mouse obesity phenotype — the gonadal fat mass (GFM), and evidences from previous studies strongly support the functional relevance between the module genes and obesity. The analysis of eQTL modules may provide important insights into the functional components underlying complex phenotypes.

Results

Formulation of the Module Detection Problem and Simulation Results

We exploited a network approach to systematically analyze large numbers of modulatory relations between genetic loci and gene expression traits. A module in an eQTL network is defined as a set of highly connected nodes — eQTLs and genes whose expression levels are regulated by some or all of the eQTLs. Only eQTLs located on different chromosomes are allowed to be included in a module to avoid trivial results caused by the linkage

between markers. A conceptual representation of eQTL module is shown in Figure 1. Module detection in an eQTL network can be formulated as an optimization problem: searching for a set of $m+n$ nodes that maximizes the objective function $Q(m, n, k) = k/(m \times n)$, where m is the number of eQTLs, n is the number of target genes and k is the number of edges between them. In this bipartite network, genes can be connected to QTLs, but there is no edge between genes and between QTLs. The maximum number of edges between n genes and m QTLs is $m \times n$, therefore Q is a value between zero and one. The objective function $Q(m, n, k)$ is a measurement of the connection density of a module. For a set of completely connected nodes, $Q = 1$; for a set of unconnected nodes, $Q = 0$. In this work, a module must have a Q value of 0.66 or above. Intuitively, this density criterion requires that on average each gene node are connected to about 2/3 or more of the QTL nodes and *vice versa*. Besides this density criterion, a module must also be statistically significant, which means the module should be highly unlikely to arise by chance in a randomized network with the same numbers of nodes and edges and the same degree distribution. The details of the module detection method are described in Materials and Methods.

A simulation study was performed to assess the performance of the module detection method. We generated random bipartite networks with prescribed modules and used normalized mutual information (NMI) [16] to evaluate the consistency between the prescribed modules and the modules identified by the search method. NMI is a robust performance indictor based on the confusion matrix [16]. The rows of the confusion matrix correspond to the prescribed modules, and the columns correspond to the identified modules. The confusion matrix contains the number of overlapped nodes between the prescribed modules and the identified module. If the identified modules completely match the prescribed modules, NMI takes the maximum value of 1.0; if the identified modules are unrelated to the simulated module, NMI becomes 0. The simulated eQTL networks consisted of 1200–1500 nodes and 3000–3500 edges, and contained 10 modules with 2–3 eQTL nodes and 20–150 gene nodes (typical sizes of the modules found in this work). Five independent simulation runs were performed with each of the following module homogeneity values: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9 (Figure 2). We then used our module detection algorithm to identify modules in the simulated networks. The details of the simulation procedure are described in Materials and Methods. The module homogeneity (p) controls the formation of the modular structures of the simulated network. For $p = 1$, the simulated network has a clear-cut modular structure. For $p = 0$, the prescribed modules become random partitions of the simulated network and therefore the network has no modular structure at all. The module detection algorithm is expected to be able to identify the prescribed modules correctly when p is high, while no module can be identified by any algorithm when p is too low. Our module detection method performed reasonably well with a NMI value above 0.8 when the module homogeneity was higher than 0.6, and the NMI value was very close to its maximum value of 1.0 when the module homogeneity is higher than 0.9. The NMI dropped quickly when the module homogeneity was below 0.5. This is because the modular structure became much fuzzier with such low module homogeneity values. For example, at a module homogeneity value of 0.5, on average only half of the edges connected to the nodes of a module come from members of the same module and the other half of the connections are randomly connected to nodes outside the module.

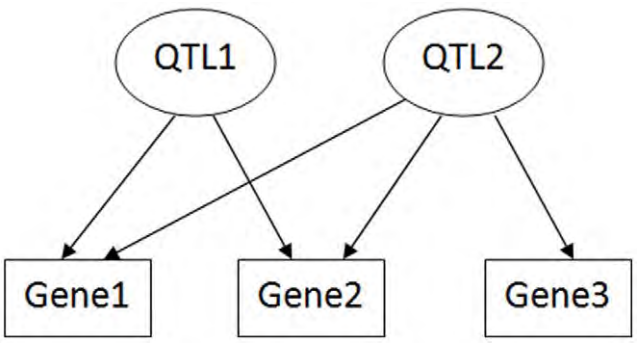


Figure 1. A conceptual representation of eQTL module. This module contains two eQTLs and three genes. The Q value of this module is 5/6.
doi:10.1371/journal.pone.0014313.g001

Expression QTL Network and Modules

The yeast eQTL network is a connected graph of 493 eQTL nodes, 4583 gene nodes, and 33,584 edges. The median degrees for the eQTL nodes and gene nodes are 25 and 7 respectively. In the yeast network, we identified 65 modules (Table S1). The number of eQTLs in each module ranges from 2 to 3, and the number of target genes ranges from 4 to 276. These modules contain 1756 unique genes, covering 38.3% of the genes in the yeast eQTL network. Three identified modules and their neighboring gene nodes in the yeast eQTL network are displayed in Figure 3. The mouse liver eQTL network is a connected graph of 408 eQTL nodes, 4086 gene nodes, and 11,458 edges. The median degrees for the eQTL nodes and gene nodes are 15 and 2 respectively. In the mouse liver network, we identified 98 modules (Table S2). The number of eQTLs in each module ranges from 2 to 4, and the number of target genes ranges from 4 to 84. These modules contain 989 unique genes, covering 24.2% of the genes in the mouse eQTL network. The size distributions of the yeast and mouse modules are shown in Figure 4. We found that these modules were highly unlikely to occur simply by chance in randomly rewired networks with the same number of nodes and edges and same degree distribution (P-value $<10^{-4}$). Therefore statistically significant modular structures exist in these eQTL networks. The modular structures of genotype-phenotype map has also been observed in some classical multiple-trait association studies [17,18].

Functional coherence of module genes

We used the Ontologizer software [19] to assess the enrichment of GO terms in each module. Ontologizer uses Parent-Child Analysis, which takes the structure of the GO hierarchy and parent-child relations into consideration when it performs the enrichment analysis. The Westfall-Young-Single-Step method [20] was used for multiple testing correction. A total of 42 yeast modules and 21 mouse modules were associated with at least one GO term at the significance level of $P<0.05$ (Tables S3 and S4).

Some modules were associated with common GO terms. For example, yeast module 63 and 64 were associated with 8 common GO terms (e.g. organelle lumen, ribosome biogenesis and assembly), and yeast module 45 and 61 were associated with 25 common GO terms (Table S3). They were identified as separate modules in the eQTL network, however there might be moderate but genuine links connecting them. These links are the weaker associations between gene expression traits of one module and eQTLs of another module, which did not pass the significance test used in eQTL mapping. We added the moderate links (with P-values <0.01 but ≥ 0.001) to the yeast eQTL

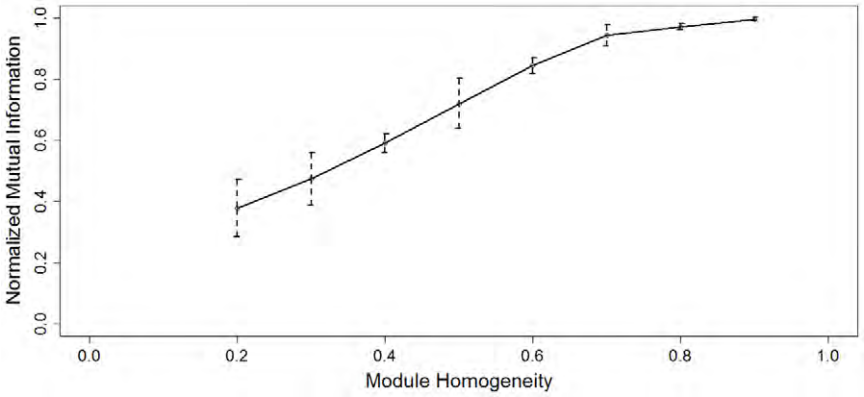


Figure 2. The performance of module detection algorithm as a function of module homogeneity. The error bars mark the interval of minus and plus one standard deviation.
doi:10.1371/journal.pone.0014313.g002

network to test if the distribution of the moderate links would suggest potential relations between the modules. We randomly rewired the moderate links in the eQTL network and counted the number of moderate links that connected each pair of modules. For each pair of modules, the maximum number of moderate links from 1000 such randomly rewired networks was compared to the number of moderate links bridging the two modules in the original network. We then sorted the module pairs by the ratio of these two numbers (original vs. rewired maximum) in a descending order. The top 20 (1%) yeast module pairs are listed in Table S5. Among the 2080 possible yeast module pairs, modules 45 and 61 ranked 18th with a ratio of 4.6, and modules 63 and 64 ranked 19th with a ratio of 4.5. There were many more (4.6 and 4.5 fold respectively) moderate links bridging these module pairs in the original eQTL network than that expected by chance in the randomly rewired networks. Other top ranked

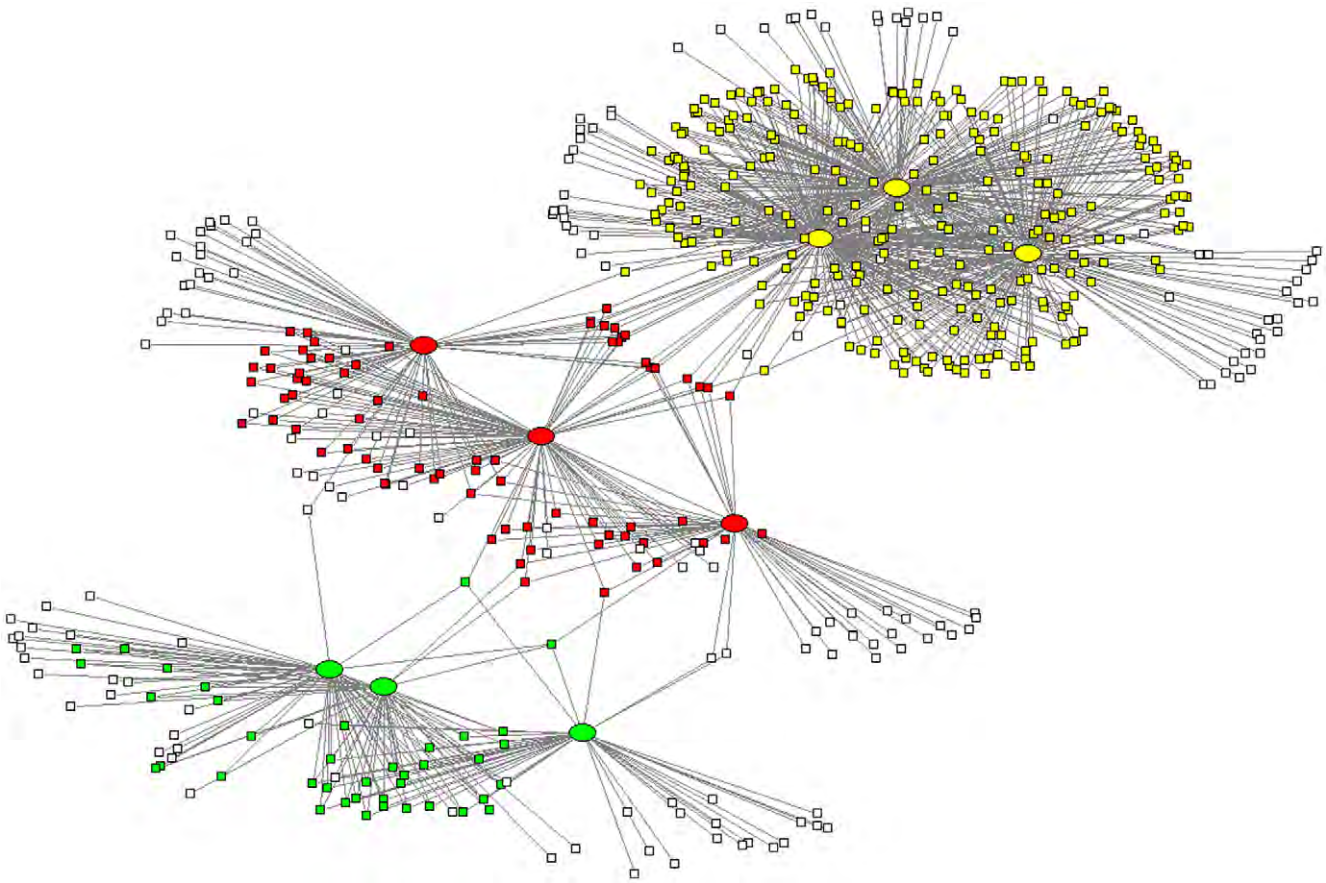


Figure 3. Three modules in the yeast eQTL network. The ellipses represents eQTLs, squares represent genes. White squares represent genes that do not belong to the three modules. Green: Module 48; Yellow: Module 64; Red: Module 55.
doi:10.1371/journal.pone.0014313.g003

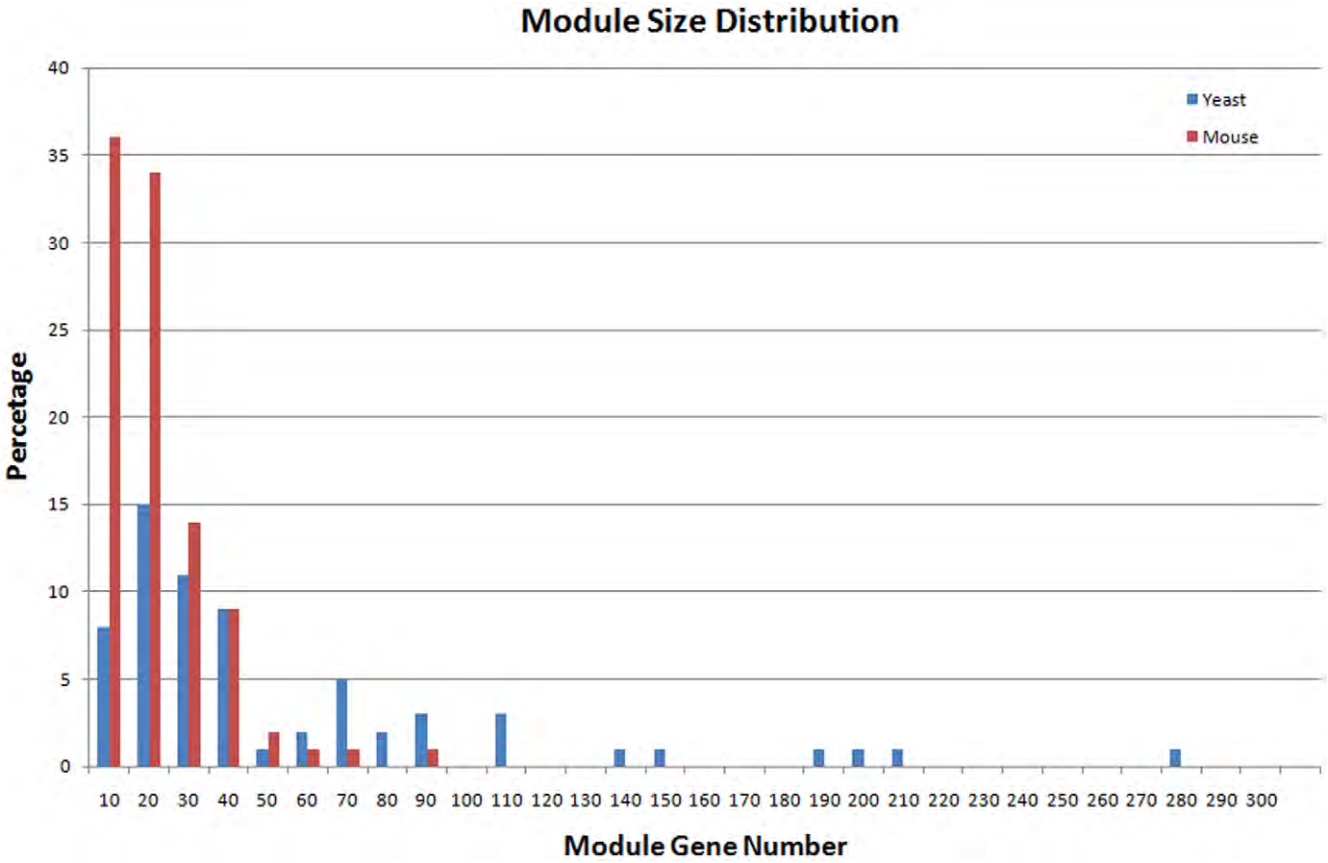


Figure 4. The size distributions of the yeast and mouse modules.
doi:10.1371/journal.pone.0014313.g004

module pairs that share common GO terms include: Module 46 and 64, Module 26 and 64, and Module 55 and 60. The non-random distribution of the moderate links may help us to identify modules that are more likely being functionally related.

Linking eQTL modules to physiological phenotypes

One major goal of systems genetics is to identify gene expression modules underlying higher-order phenotypes. Recently, Nogami et al. [21] measured more than 500 yeast morphologic phenotypes and mapped 7 significant QTLs (false discovery rate [FDR] = 0.05) (Table 2 and Table S4 of [21]). We assessed the genetic overlap between these 7 morphologic QTLs and the yeast eQTL modules we identified. We found that QTLs on three chromosomes were shared by morphologic phenotypes and the modules (Table 1). The morphologic phenotypes can be classified into six categories, each representing an aspect of cellular morphology (Table 2 of [21]). Phenotypes of same category were usually mapped to QTLs on same chromosome. But there was a surprising exception where the phenotypes concerning DNA region size, position, and shape were mapped to two unlinked loci on Chromosome 14 and 15, respectively [21]. The eQTL module analysis may provide a possible explanation to the exception. The modules with eQTLs on Chromosome 14 and 15 were associated with different GO terms. Three modules (28, 45, and 61) with eQTLs on Chromosome 14 were associated with protein metabolism while three modules (7, 9, and 51) with the QTLs on Chromosome 15 were associated with mitochondrial oxidative phosphorylation and energy generation. This indicates different molecular pathways may underlie the phenotypes mapped to chromosome 14 and those mapped to chromosome 15 though they all belong to same category.

We also analyzed the physiological relevance of the mouse liver eQTL modules. The obesity phenotype gonadal fat mass (GFM) was genetically dissected, and five "clinical" QTLs (cQTLs) regulating this phenotype were mapped in a previous study (Table 2 of [22]). We analyzed the overlaps between the module QTLs and these five cQTLs. Three modules (50, 74, and 84) had eQTLs that overlapped with a cQTL on chromosome 19. Module 74 was of particular interest because it had another eQTL located near a cQTL on chromosome 5. The distance between the two QTL markers is about 20 Mb. This module contains three eQTLs and 21 genes, seven of which were uncharacterized expressed sequence tags (ESTs) (Figure 5). There is literature evidence for seven of the module genes (i.e. 50% of the genes in this module with known functions) being related to obesity. *Lcat* (lecithin cholesterol acyltransferase) is involved in lipid metabolism which affects the GFM trait [23]. Other module genes related to lipid metabolism and obesity include *Anxa5* (annexin A5) [24], *Ccna2* (cyclin A2) [25], *Ces5* (carboxylesterase 5) [26], *Cyp2c38* (cytochrome P450, family 2, subfamily c, polypeptide 38) [27], *Setd8* (SET domain containing 8) [28], and *Slc16a11* (monocarboxylic acid transporters, member 11) [29]. Thus, literature evidence supports the association between GFM trait and the eQTL module.

Discussion

In this work we exploited a network approach to systematically analyze large numbers of modulatory relations between genetic

Table 1. Genetic overlap of yeast eQTL modules and morphologic phenotypes.

Phenotype category	QTL (bp)	Module ID
DNA region size, position and shape	chr14:440000-460000	21, 28, 50, 59, 61, 62, 63
	chr14: 480000-500000	24, 52, 58
	chr14: 500000-520000	13, 45
DNA region size, position and shape	chr15: 520000-540000	9
	chr15: 540000-560000	7, 21, 51
Mother cell size and shape	chr8: 60000-80000	56, 58
	chr8: 80000-100000	27, 37
	chr8: 100000-120000	43

doi:10.1371/journal.pone.0014313.t001

loci and gene expression traits. Like many other biological networks eQTL networks have evolved functional modules. Such modular structures may confer selective advantage by allowing the optimization of gene expression within each module and therefore minimizing the impact of genetic variants outside the module.

Recently Zhang et al. [15] used a Bayesian partition method to identify eQTL modules from the same yeast dataset used in this work. They identified 20 yeast modules with one eQTL and 9 modules with two eQTLs. In this work we are interested in detecting eQTL modules with complex genetic architectures. Therefore we focused on modules with at least two QTLs in different chromosomes, and identified 21 yeast modules with two eQTLs and 44 yeast modules with three eQTLs. The Bayesian partition method [15] essentially performs eQTL mapping and module identification simultaneously. Our module detection method takes the eQTL mapping results as the input and can be used with any eQTL mapping method; therefore it provides the flexibility to reanalyze the eQTL network when new algorithms for eQTL mapping become available.

Epistasis is a higher-order genetic interaction that go beyond the pair-wise regulatory relations between a QTL and a trait. To test the epistatic effects within the eQTL modules, we employed a regression based model selection approach to find the best eQTL model for each module gene. The expression values of each

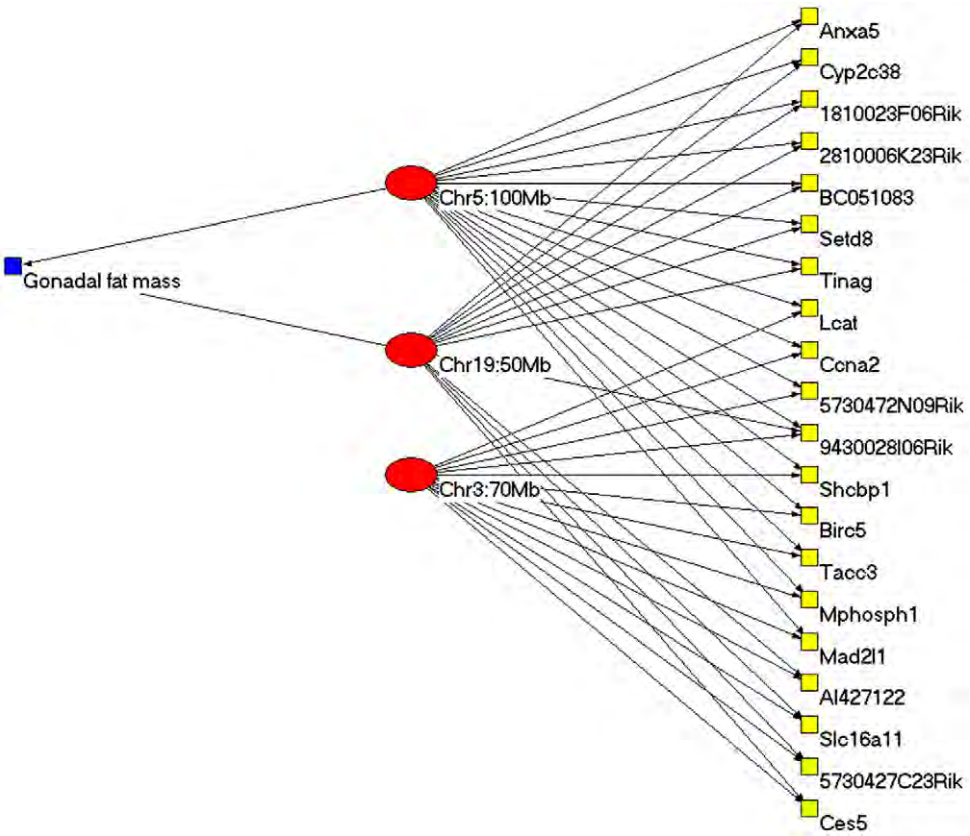


Figure 5. Genetic overlaps between Mouse gonadal fat mass (GFM) trait and module 74. Red ellipses represent QTLs, yellow squares represent genes and the blue square represents the GFM trait.
doi:10.1371/journal.pone.0014313.g005

module gene were regressed on the genotypes of the module eQTLs with or without interaction terms. Including interaction terms in an eQTL model may improve the model fit but also increases the model complexity. We used the standard Akaike Information Criterion to select the eQTL model with the best tradeoff between goodness-of-fit and model complexity [30]. Then for each module, we computed the proportion of module genes that could be best modeled by including the epistatic interactions. We found that this proportion ranged from 0% to 59% in the yeast modules with a median of 19%. Only four modules do not include epistatic QTLs. These results have revealed the genetic complexity of the eQTL modules.

We compared the eQTL modules of the two organisms and found that 21 (32.3%) and 44 (67.7%) yeast modules have two and three eQTLs respectively, while 12 (12.2%) and 85 (86.7%) mouse modules have two and three eQTLs respectively, and one mouse module has four eQTLs. The median module gene numbers for the yeast and the mouse modules are 27 and 14 respectively. The higher percentage of mouse modules with three eQTLs and the lower number of genes in mouse modules (Fig. 4) indicates that the regulation of gene expression in mouse is more genetically complex than that in yeast.

Materials and Methods

Construction of eQTL networks

We used two data sets, a yeast dataset [31,32] and a mouse liver dataset [22], to construct the eQTL networks. The yeast data set contained genotype data of 2957 markers and gene expression data of 6216 open reading frames in 112 F1 segregants that were generated by crossing the BY4716 strain with the RM11-1a strain. Linkage analysis was performed using the Wilcoxon test, and statistical significance was estimated by permutations [31]. Significant linkage results with P-values <0.001 were used to construct the eQTL network. We divided the yeast genome into bins of 20 Kb and mapped QTLs onto them. In an eQTL network, two types of nodes were used to represent eQTLs and gene expression traits respectively, and edges represent the modulatory relations between the QTLs and gene expression traits. Gene expression traits mapped to only one eQTL were not included in the network because such nodes would not belong to any module.

The mouse liver data set contained expression data of 23,574 mouse transcripts in the livers of 334 F2 mice generated by crossing the C57BL/6J ApoE$^{-/-}$ strain with the C3H/HeJ ApoE$^{-/-}$ strain [22]. Using this data set, Wang et al. [22] mapped suggestive and significant eQTLs (Table S1 of [22]). Again, we used $P<0.001$ as the cutoff value to construct the eQTL network. We divided the mouse genome into bins of 5 Mb and mapped QTLs onto them. The QTLs modulating the GFM trait of these mice were also mapped by Wang et al. (Table 2 of [22]).

Module detection

We employed a two-step search algorithm: in the first step we tried to find as many seed modules as possible and in the second step we merged overlapping seed modules. We searched for seed modules within a range of m (the number of eQTL nodes) from 2 to 6 and n (the number of gene nodes) from 4 to 14. For each combination of m and n, we started with a randomly picked, connected set of m eQTL nodes and n gene nodes. In each following step, one node in the current set was randomly selected and an attempt was made to replace it with a randomly picked node that does not belong to the current set but is connected to the current set by one or more edges. At the end of every 25 steps, one node in the current set was replaced with a node that had no connection to the current set to avoid getting stuck in local maxima. Changes were accepted or rejected according to the Metropolis criteria [33,34], i.e. a move was accepted with a probability of the smaller of 1.0 or $\exp[(Q_{new}-Q_{old})\times 10]$ where Q_{new} and Q_{old} were the new and old Q values. The optimization continued until $Q=1$ or 500 moves had been made. One thousand such searches with different random starts were performed and all identified seed modules (i.e., sets of connected node with a $Q \geq 0.66$ and a P-value $<10^{-4}$) were recorded. These seed modules were then merged iteratively. Each time two overlapping seed modules were merged if and only if the resulting module still had a $Q \geq 0.66$ and a P-value $<10^{-4}$. This process continued until no further merging was possible. The P-values for modules of different sizes (i.e. each combination of m and n) were estimated by random rewiring. One thousand networks were generated by randomly rewiring the edges of the original eQTL network, while keeping the edge degree of each node unchanged. The rewiring scheme is adopted from [35]. Two edges (A–B and C–D) are randomly selected and then rewired such that the new edges are A–D and B–C, provided neither of these new edges exists in the current network. This rewiring scheme is equivalent to randomly switching pairs of 0 and 1 in the rows of the adjacency matrix while keeping the raw and column margins unchanged. We then applied the module detection algorithm to these randomized networks to estimate the statistical significance of the Q value for each combination of m and n. One thousand independent searches with different random starts were performed for each randomized network.

Simulation

In order to access the performance of our model detection algorithm, we generated random networks with prescribed modular structure and then used our method to identify the predefined modules. We adopted the module simulation method for bipartite network as described in [36] with minor changes to accommodate the module density criterion (Q value) used in this work. We first predefined the module membership for all the eQTL and gene nodes being considered. We also predefined N_i, the number of gene nodes within the i-th module. For each eQTL node, we connected it to N_i gene nodes: with probability p, a gene node randomly selected from the same module was connected to the eQTL; otherwise a gene node randomly selected from the whole gene node set was connected to the eQTL. The parameter p controls the degree of homogeneity of a module and hence is called module homogeneity. If a module generated this way did not satisfy our module density criterion ($Q \geq 0.66$), we extracted a subset of nodes from the module that met this criterion as the final module. The normalized mutual information (NMI) was used to assess the performance of the search algorithm. Given a confusion matrix in which rows are prescribed modules and columns are detected modules, NMI is defined as

$$\frac{-2\sum_{M1}\sum_{M2} N_{ij}\times\log\left(\frac{N_{ij}\times N}{N_{i.}\times N_{.j}}\right)}{\sum_{M1} N_{i.}\times\log\left(\frac{N_{i.}}{N}\right)+\sum_{M2} N_{.j}\times\log\left(\frac{N_{.j}}{N}\right)},$$

where N_{ij} is an element of the confusion matrix specifying the number of overlapped nodes between the i-th prescribed module and the j-th detected module. $N_{i.}$ and $N_{.j}$ are the row means and column means respectively, and M_1 and M_2 are the number of prescribed and detected modules [16].

Supporting Information

Table S1 List of Yeast Modules
Found at: doi:10.1371/journal.pone.0014313.s001 (0.57 MB XLS)

Table S2 List of Mouse Liver Modules
Found at: doi:10.1371/journal.pone.0014313.s002 (0.19 MB XLS)

Table S3 Gene Ontology Analysis of Yeast Modules
Found at: doi:10.1371/journal.pone.0014313.s003 (0.06 MB XLS)

Table S4 Gene Ontology Analysis of Mouse Liver Modules
Found at: doi:10.1371/journal.pone.0014313.s004 (0.03 MB XLS)

Table S5 Yeast module pairs with significant numbers of moderate between-module links
Found at: doi:10.1371/journal.pone.0014313.s005 (0.02 MB XLS)

Acknowledgments

We thank Dr. Rachel Brem for providing the yeast genotype data and eQTL mapping results. We thank Dr. Gael Yvert for providing yeast morphological trait data.

Author Contributions

Conceived and designed the experiments: LB YC. Performed the experiments: LB. Analyzed the data: LB. Contributed reagents/materials/analysis tools: XX. Wrote the paper: LB YC.

References

1. Rockman MV, Kruglyak L (2006) Genetics of global gene expression. Nat Rev Genet 7: 862–872.
2. Quigley D, Balmain A (2009) Systems genetics analysis of cancer susceptibility: from mouse models to humans. Nat Rev Genet 10: 651–657.
3. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, et al. (2009) Systems genetics of complex traits in Drosophila melanogaster. Nat Genet 41: 299–307.
4. Schadt EE (2009) Molecular networks as sensors and drivers of common human diseases. Nature 461: 218–223.
5. Rockman MV (2008) Reverse engineering the genotype-phenotype map with natural genetic variation. Nature 456: 738.
6. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, et al. (2008) Genetics of gene expression and its effect on disease. Nature 452: 423–428.
7. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, et al. (2008) Variations in DNA elucidate molecular networks that cause disease. Nature 452: 429–435.
8. Bao L, Peirce JL, Zhou M, Li H, Goldowitz D, et al. (2007) An integrative genomics strategy for systematic characterization of genetic loci modulating phenotypes. Hum Mol Genet 16: 1381–1390.
9. Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet 17: 388–391.
10. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297–302.
11. Barabasi A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 101–113.
12. Rives AW, Galitski T (2003) Modular organization of cellular networks. Proceedings of the National Academy of Sciences of the United States of America 100: 1128–1133.
13. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, et al. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet 34: 166–176.
14. Li H, Chen H, Bao L, Manly KF, Chesler EJ, et al. (2006) Integrative Genetic Analysis of Transcription Modules: Towards Filling the Gap between Genetic Loci and Inherited Traits. Hum Mol Genet 15: 481–492.
15. Zhang W, Zhu J, Schadt EE, Liu JS (2010) A Bayesian Partition Method for Detecting Pleiotropic and Epistatic eQTL Modules. PLoS Comput Biol 6: e1000642.
16. Danon L, Diaz-Guilera A, Duch J, Arenas A (2005) Comparing community structure identification. Journal of Statistical Mechanics-Theory and Experiment. 10 p.
17. Klingenberg CP, Leamy LJ, Cheverud JM (2004) Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. Genetics 166: 1909–1921.
18. Mezey JG, Cheverud JM, Wagner GP (2000) Is the genotype-phenotype map modular? A statistical approach using mouse quantitative trait loci data. Genetics 156: 305–311.
19. Robinson PN, Wollstein A, Bohme U, Beattie B (2004) Ontologizing gene-expression microarray data: characterizing clusters with Gene Ontology. Bioinformatics 20: 979–981.
20. Westfall PH, Zaykin DV, Young SS (2002) Multiple tests for genetic effects in association studies. Methods Mol Biol 184: 143–168.
21. Nogami S, Ohya Y, Yvert G (2007) Genetic complexity and quantitative trait loci mapping of yeast morphological traits. PLoS Genet 3: e31.
22. Wang S, Yehya N, Schadt EE, Wang H, Drake TA, et al. (2006) Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. PLoS Genet 2: e15.
23. Greaves KA, Going SB, Fernandez ML, Milliken LA, Lohman TG, et al. (2003) Cholesteryl ester transfer protein and lecithin:cholesterol acyltransferase activities in hispanic and anglo postmenopausal women: associations with total and regional body fat. Metabolism 52: 282–289.
24. van Tits L, de Graaf J, Toenhake H, van Heerde W, Stalenhoef A (2005) C-Reactive Protein and Annexin A5 Bind to Distinct Sites of Negatively Charged Phospholipids Present in Oxidized Low-Density Lipoprotein. Arterioscler Thromb Vasc Biol 25: 717–722.
25. Laudes M, Bilkovski R, Oberhauser F, Droste A, Gomolka M, et al. (2008) Transcription factor FBI-1 acts as a dual regulator in adipogenesis by coordinated regulation of cyclin-A and E2F-4. Journal of Molecular Medicine 86: 597–608.
26. Soni KG, Lehner R, Metalnikov P, O'Donnell P, Semache M, et al. (2004) Carboxylesterase 3 (EC 3.1.1.1) is a major adipocyte lipase. J Biol Chem 279: 40683–40689.
27. Kudo T, Shimada T, Toda T, Igeta S, Suzuki W, et al. (2009) Altered expression of CYP in TSOD mice: a model of type 2 diabetes and obesity. Xenobiotica 39: 889–902.
28. Wakabayashi K-i, Okamura M, Tsutsumi S, Nishikawa NS, Tanaka T, et al. (2009) The Peroxisome Proliferator-Activated Receptor {gamma}/Retinoid X Receptor {alpha} Heterodimer Targets the Histone Modification Enzyme PR-Set7/Setd8 Gene and Regulates Adipogenesis through a Positive Feedback Loop. Mol Cell Biol 29: 3544–3555.
29. Merezhinskaya N, Fishbein WN (2009) Monocarboxylate transporters: past, present, and future. Histol Histopathol 24: 243–264.
30. Akaike H (1974) A New Look at the Statistical Model Identification. IEEE Trans Autom Control 19: 716–723.
31. Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci USA 102: 1572–1577.
32. Brem RB, Storey JD, Whittle J, Kruglyak L (2005) Genetic interactions between polymorphisms that affect gene expression in yeast. Nature 436: 701–703.
33. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21: 1087–1092.
34. Spirin V, Mirny LA (2003) Protein complexes and functional modules in molecular networks. Proceedings of the National Academy of Sciences of the United States of America 100: 12123–12128.
35. Maslov S, Sneppen K (2002) Specificity and Stability in Topology of Protein Networks. Science 296: 910–913.
36. Guimera R, Sales-Pardo M, Amaral LAN (2007) Module identification in bipartite and directed networks. Physical Review E 76: 8.

Comparison of variant detection using whole genome sequencing of the DBA/2J mouse strain

Xusheng Wang[1], Megan K. Mulligan[1], Khyobeni Mozhui[1], Lu Lu[1], Zugen Chen[2], Stanley F. Nelson[2], William L. Taylor[1], Robert W. Williams[1]

1. University of Tennessee Health Science Center, Memphis TN 38163
2. University of California, Los Angeles CA 90095

Introduction

Next Generation Sequencing (NGS) technology is now widely used to detect sequence and structural variants. Here we ask how coverage depth and platform affect the rate of detection of major classes of variants.

Results

We generated ~58X coverage for the DBA/2J genome using Illumina GAII and HiSeq2000 systems and ~42X using ABI SOLiD. We used six paired end libraries with insert lengths from 200 to 4,000 bp, and aligned against the C57BL/6J genome. We detected 4.16 million SNPs using Illumina and 4.09 M using SOLiD, of which 3.38 M SNPs were common. Unshared SNPs were validated 93% of the time by resequencing with a total yield (~99.7% true) of 4.87 M between strains. Each platform detects an large cohort of unique SNPs (15% per platform). We detected 0.56 M and 0.22 M indels using Illumina and SOLiD, respectively. Only 0.11 M were common. Finally, we identified 5,600 and 8,800 structural deletions using Illumina and SOLiD or which only 1,223 were common. An analysis of Illumina subsamples at 10, 20, 30, 40, and 55X demonstrates that platforms can rapidly reach a premature SNP detection asymptote that cannot be overcome simply by higher coverage. Indel and CNV detection is more challenging and even at 100–120X we appear to be far from a "full disclosure" on sequence variants even in these two strains. As is true for SNPs, each platform detects unique and genuine subsets of indels and CNVs. This incompleteness problem is compounded by the fact that assembly is biased by using a C57BL/6J scaffold and limitations of approaches used to detect variants.

Conclusion

Sequencing data from multiple platforms with high coverage will be necessary for the next few years to extract the majority of variants, particularly large structural variants.

Methods

Genomic DNA was isolated from liver of two DBA/2J females. We sequenced DNA using ABI SOLiD and Illumina GA2/Hiseq2000 platforms. GA2/Hiseq2000 reads were paired-end 100 bp-long whereas SOLiD reads were mate-paired and 25 and 50 bp-long. We also generated an amplification-free Illumina sequencing-library, which enables us much improve de novo assembly of GC-biased regions and thereby detect more variants.

●***Mapping and alignment***
Reads were mapped to the C57BL/6J reference mouse genome (mm9) using MAQ and the SOLiD Corona Lite Pipeline for GA2 and SOLiD reads, respectively.

●***Single nucleotide polymorphisms (SNPs) calling***
SNPs from GA2/Hiseq2000 data sets were identified by MAQ SNP calling with a minimal read depth of 3 and a minimum consensus quality scores of 30 (P < 10E-3). SNPs from SOLiD data sets were also detected with a minimal read depth of 3 and a stringent confidence value of 0.5.

●***Insertions and deletions***
Indels were detected by MAQs and ABI Corona pipeline for GA2/ Hiseq2000 and SOLiD data sets, respectively. Large deletions (1 bp – 10 kb) and medium-sized insertions (1 – 86 bp) were detected using GA2/HiSeq2000 reads. Smaller insertions (1 – 11 bp) and deletions (1 - 36 bp) were detected using 50 bp SOLiD reads .

●***Copy number variants (CNVs)***
Copy number variants were analyzed using an event-wise testing (EWT) method that is based on read depth and significance testing A common method was used for both platforms.

●***Variants detection at different levels of coverage***
We generated 55X coverage using Illumina. We divided data into 10, 20, 30, 40 and 55X coverage—lane by lane—to detect SNPs, indels, and CNVs as described above. The sum of all variants detected using the combined 100X coverage was used as a reference against which subsamples were evaluated.

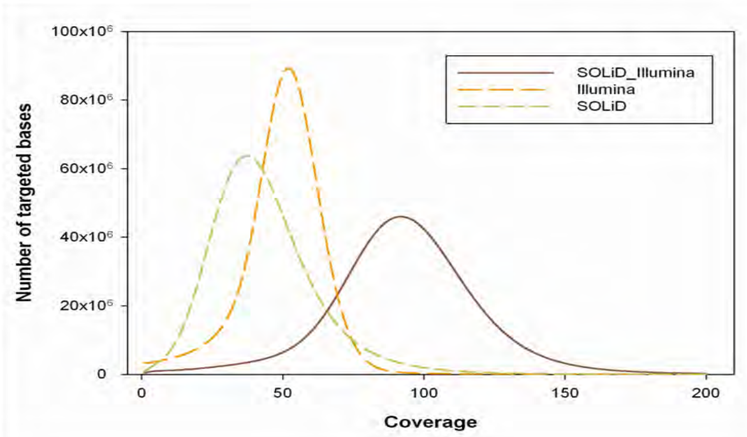
Summary of sequencing data for the DBA/2J genome

Summary of sequencing data

Platform	ABI SOLiD	Illumina GA2
Run number	15	3
Read Length	25 bp / 50 bp	100 bp
Clone insert size	1k to 4k	200 bp
Mapped reads	3.06 billion	1.43 billion
Mapped nucleotides	108 billion	153 billion
Coverage	~ 42X	~58X

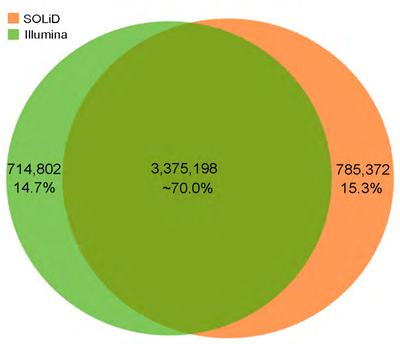
Assumed C57BL/6J reference genome size is 2.6 Gb. Read depth was determined by calculating all mapped nucleotides divided by the size of mouse genome.

Distribution of coverage



Variants comparison between SOLiD and Illumina

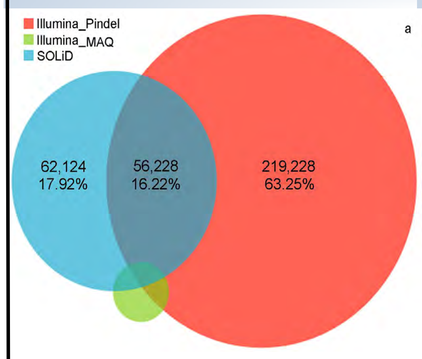
SNPs



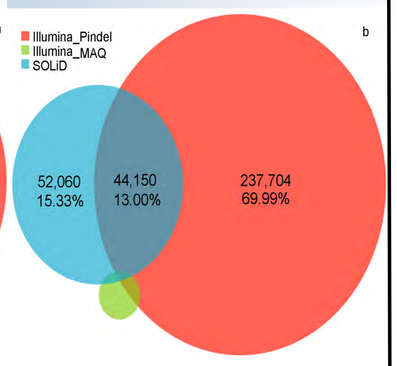
Genomes of DBA/2J and C57BL/6J differ from each other at 4.875 million SNPs

Platform	N SNPs	False Positive
SOLiD	4,160,570	0.44%
Illumina	4,090,000	0.65%

Insertions



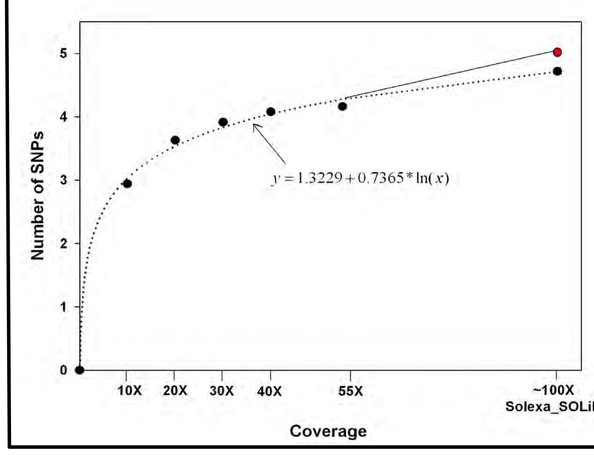
Deletions



More than half of indels were detected exclusively by Illumina , whereas ~15% were detected exclusively by SOLiD.

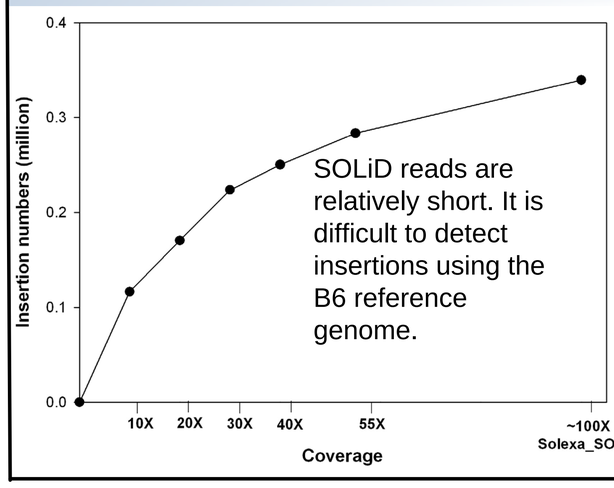
Variants detection at different levels of coverage

SNPs



SNP summary:
1. A total number of 4.875 million SNPs were detected by combining Illumina and SOLiD data (red dot).
2. We estimate that 100X Illumina data would detect 4.71 million SNPs (black dot).
3. 98% at SNPs are detected at 40X.

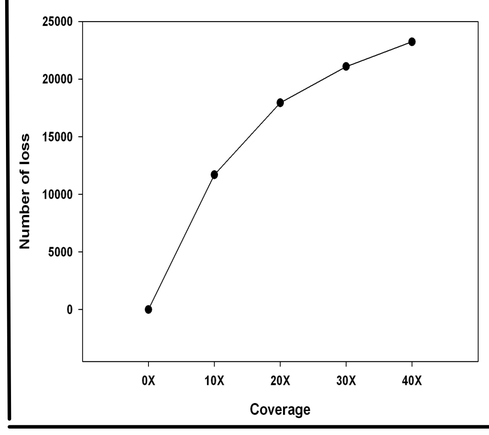
Insertions



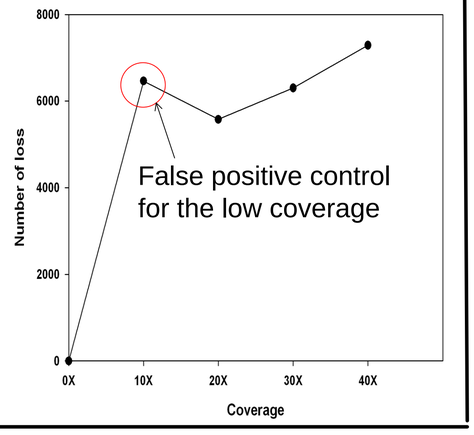
SOLiD reads are relatively short. It is difficult to detect insertions using the B6 reference genome.

CNVs

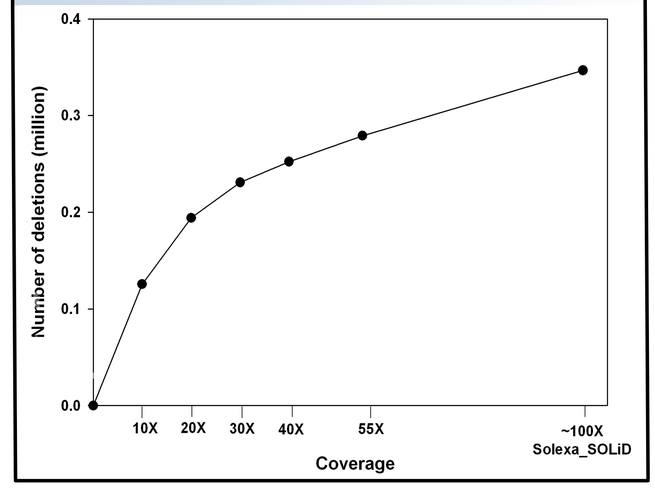
Loss



Gain



Deletions



Acknowledgments

This work supported by the UT Center for Integrative and Translational Genomics and grants from NIAAA Integrative Neuroscience Initiative on Alcoholism (U01AA13499, U24AA13513 and U01AA014425).