5-12-2017

Preprocessing Strategies for Multiplex Bead Assay Data for Use in Quantitative Trait Loci Analysis

Circe McDonald

PREPROCESSING STRATEGIES FOR MULTIPLEX BEAD ASSAY DATA FOR USE IN QUANTITATIVE TRAIT LOCI ANALYSIS

By

CIRCE ELIZABETH MCDONALD

Preprocessing Strategies for Multiplex Bead Assay Data for Use in Quantitative Trait Loci Analysis

ABSTRACT

INTRODUCTION:

Host genetic variants are known to impact infectious disease susceptibility and outcomes. However, the genes underlying these impacts are not well characterized. Multiplex bead assays (MBA) provide an affordable and rapid means of large scale screening for multiple phenotypic measures of immune response. Transformation and normalization approaches for MBA data have not been agreed upon, especially concerning screening applications.

AIM:

To compare preprocessing techniques in improving validity of quantitative loci trait analyses which utilize MBA phenotypic data with high levels of predictor technical variability using experimental data.

METHODS:

This research uses primary dendritic cells derived from a set of sixty-one genetically diverse mouse strains to study activation response of an antiviral pathway (RIG-I). Primary outcomes were IFNα and IFNβ secretion following RIG-I agonist treatment. Multiple transformation and normalization approaches were used to estimate true IFNα and IFNβ responses. Evaluation criteria included three quantitative measures (tail length, kurtosis, skewness) and three qualitative measures (QQ-plot, Bland-Altman plot, Mean-SD plot).

RESULTS:

Most qualitative measures and quantitative measures found log transformation with quantile normalization was most appropriate for normalizing data and reducing technical variability

between batches and replicates. Unfortunately, no statistically significant ($\alpha = 0.90$) loci of interest were identified with this normalized data.

DISCUSSION:

The data used to test these methods had notable limitations, mainly only two phenotypic markers and dramatic variability in both technical and biological replicates. While normalization and transformation techniques did ameliorate these issues, additional approaches such as mixed effects modeling may be able to further improve these types of analysis.

PREPROCESSING STRATEGIES FOR MULTIPLEX BEAD ASSAY DATA FOR USE IN QUANTITATIVE TRAIT LOCI ANALYSIS

By

CIRCE ELIZABETH MCDONALD

B.S., DOMINICAN UNIVERSITY OF CALIFORNIA

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

APPROVAL PAGE

PREPROCESSING STRATEGIES FOR MULTIPLEX BEAD ASSAY DATA FOR USE IN QUANTITATIVE TRAIT LOCI ANALYSIS

By

CIRCE ELIZABETH MCDONALD

Approved:

_____
Ruiyan Luo, Committee Chair

_____
Mehul Suthar, Committee Member

_____
Date

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

_____

Circe McDonald

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

There is an increasing awareness of and appreciation for the role that genetic variants play in host susceptibility and disease outcomes in the context of infectious disease[1,2]. Improved understanding in this area has the potential to inform epidemiology studies in endemic and outbreak scenarios as well as proactive interventions against emerging infectious diseases[3]. Further, understanding the mechanisms underlying these variations may yield additional insight into the biological processes influencing disease lifecycle.

Viral diseases provide an excellent means of studying the impact of host genetics on the course of infection due to the diversity of host immune response, highly characterized viral genomes, and availability of animal models. Particularly, viruses in the genus *Flavivirus,* a group of single-stranded, vector born RNA viruses, are of interest due to their unique interactions with host immune systems. Two members of this genus are of particular public health interest in the United States, West Nile virus (WNV) and Zika virus (ZikV). WNV accounted for 2,175 deaths in the US in 2015 and is actively transmitted in a majority of states[4]. In early 2016, ZikV spread to Puerto Rico and Florida, with over 15,000 locally acquired infections, 18 cases of birth defects, and 6 cases of pregnancy loss reported far[5]. Both these viruses have shown significant variation in disease outcomes in *in vitro* models[6], animal models[7], and human subjects[8,9]. While the neurological effects of WNV and the birth defects of ZikV can be life threatening, the majority of infected individuals are asymptomatic in the presence of viral replication[8,9]. Differences between hosts which underlie these divergent outcomes are still poorly understood though antiviral pathways involved in response, such as RIG-I and MDA5, have been more extensively studied.

We hypothesized that host genetic variation can partially explain differences in cellular response to WNV and ZikV infections based in the RIG-I signaling pathway. This research used primary dendritic cells derived from a set of sixty-one genetically diverse and characterized mouse strains, known as Collaborative Cross (CC) strains, to study cellular response to RIG-I antagonist treatment as well as viral infection (WNV and ZikV). This mouse population has been previously used to study influenza and was found to mirror some aspects of diverse human immune responses during viral infection[10]. Measured outcomes include two surface markers of dendritic cell activation (CD86, CD40), excretion of cytokines IFNα and IFNβ, transcription of IFNβ, viral titers, quantity of viral RNA, and viral protein expression (WNV only) all measured at 48h post infection. Phenotypic outcomes were associated with single nucleotide polymorphisms (SNPs) in CC strain genomes. Using quantitative trait loci analysis (QTL), we determined the statistical significance and robustness of these associations to identify genomic regions contributing to observed phenotype outcomes.

Preliminary analysis found IFNβ response measured by commercial multiplex bead assay (MBA) captured notable variation between strains with both RIG-I antagonist treatment and viral infection and was therefore suitable for QTL analysis. However, measurement and normalization approaches for MBA data have not been universally agreed upon, especially concerning the application of MBA for screening and association studies. Batch and plate effects were noticed with included controls and replicates. Multiple approaches have been proposed for handling MBA data, but none have been specifically tailored to QTL analysis and small numbers of analytes. Different assays capturing the same phenotypic data often employ unique pre-processing strategies, such as microarray and RT-PCR, and standard normalization

techniques cannot be assumed to be appropriate for diverse assays. Because relatively small differences were seen in screening, QTL analysis is likely to be highly sensitive to any pre-processing approach. We compared combinations of five transformation (none, log2, asinh, Box-Cox, Weighted Box-Cox) and four normalization (none, quantile, loess, reduced spline normalization) strategies to estimate true IFNα and IFNβ responses. Each strategy was tested in the normalization and transformation of data generated from previously described CC screening experiment. Evaluation criteria included three quantitative measures (tail length, kurtosis, skewness) and three qualitative measures (QQ-plot, Bland-Altman plot, Mean-SD plot). Qualitative measures were reviewed blinded by two separate raters with inter-rater and intra-rater reliability calculated.

The most promising normalization and transformation method was selected based off the results of these measures. After preliminary analysis, primary outcomes were chosen to be IFNα and IFNβ secretion following RIG-I agonist treatment. The raw data and normalized data were used in parallel in QTL analysis to demonstrate the impact that normalization had upon analysis outcome. However, only the normalized data was fully analyzed for loci of interest to avoid misinterpretation. No statistically significant ($\alpha = 0.90$) results were found. Most significant loci of interest were analyzed to determine if they suggested any genes related to viral immune response.

LITERATURE REVIEW

2.1 Public Health Implications of Flaviviruses

Family *Flaviviridae* genus *Flavivirus* includes over 70 viruses many of which are pathogenic to humans, including dengue virus, Japanese encephalitis, yellow fever, West Nile virus, and Zika virus. This family is collectively referred to as flavivurses. Flaviviruses are 40-60nm, enveloped, positive-sense single-stranded RNA viruses[11]. While each virus's life history is unique, the main route of transmission is through infected arthropods (mosquitos or ticks). Flaviviruses infect a wide range of species including mammals[12], birds[13,14], and reptiles[15] with many being zoonotic[16,17]. Symptoms from flavivirus infections vary greatly by host species, strain, and host genetics including fever, birth defects, encephalitis, and death. Given the prevalence of this viral family, spillover events and continued emergence of infectious disease are likely, especially in the context of changes in biodiversity and growing interfaces between humans and wildlife[9,18]. Further understanding how host genetics impact immune responses on the cellular level can elucidate the health impacts of this family of viruses in both humans and wildlife. This project is interested particularly in two viruses of public health importance, WNV and ZikV, discussed below.

Since being introduction to the United State in 1999, West Nile Virus (WNV) has led to over 1,900 human deaths with nearly 43,000 reported cases[4]. It is estimated that 80% of WNV infected individuals remain assymptomatic[19]. For those individuals who experience symptoms, the most common manifestations are fever and rash. Rarer and more severe symptoms occur in about 1% of the population with elderly and immunocompromised individuals at higher risk. Severe symptoms stem from neuroinvasive WNV resulting in meningitis, encephalitis, and

poliomyelitis[19] and may lead to death. It is worth noting that birds are WNV's natural host and serve as the primary reservoir for the disease with Blue Jays, House Finches, and American Crows being particularly susceptible making eradication and control efforts difficult. American Crows are estimated to have an approximately 89% mortality rate with WNV infection[20]. Five lineages are also believed to have recently emerged presenting with different morbidities and mortalities[13,21]. WNV is expected to continue to be a public health concern in the United States and globally.

Zika virus (ZikV) became a serious public health concern following the discovery of an association between viral infection in pregnant women and rick of microcephaly. An increase in microcephaly incidence in Brazil in early 2015 is suspected to have been caused by a ZikV pandemic[22]. Since the beginning of concerted monitoring by the CDC in January 2015, 5,182 cases have been reported within the United States with 222 cases believed to be due to local transmission in Florida and Texas[5]. Symptoms are generally mild to moderate and, similarly to WNV, mainly consist of rash, fever, joint paint, and conjunctivitis[9,23,24]. It is suspected that a majority of ZikV infections are asymptomatic, but thorough screening studies have yet to be publication. Exact rate of birth defects in the presence of ZikV infection is also a matter of ongoing study[25,26]. No evidence of disease due to Zika virus has been reported in other species except for non-human primates, which show mild symptoms[5]. While ZikV is no longer considered a "public health emergency of international national concern", active transmission has been established in multiple countries and the birth defects seen following infection will have long term ramifications for communities, especially within South and Central America as well as the U.S. territories of Puerto Rico, American Samoa, and the U.S. Virgin Islands.

The burden of disease for this family of viruses is difficult to calculate and estimates vary greatly[27,28]. Further, the burden varies by region often with less developed regions having both a limited surveillance structure and the highest burden[27]. While vaccines exist for yellow fever virus, Japanese encephalitis virus, and various tick-borne encephalitis viruses, WNV and dengue virus vaccines are in development but facing significant hurdles, primarily cost effectiveness given low incidence of disease and the need for a tetravalent vaccine respectively[29]. Zika virus vaccine efforts have been somewhat fast tracked with the 2015 epidemic[30], however the immune response to ZikV and potential vaccines as well as how this response varies within the diverse human population is still unknown[31].

2.2 The Innate Immune System and RIG-I Signaling Pathway

Responses to viral infection rely upon the innate immune system, a nonspecific defense mechanism which is activated by the chemical properties of an antigen. These properties are collectively referred to as pathogen-associated molecular patterns (PAMPs). PAMPs include envelope proteins, protein-lipid complexes, and viral nucleic acid and may be found on the cell surface of infected cells, in the cytoplasm, and within endosomes. These PAMPs activate pathogen recognition receptors (PRRs) which cause a signaling cascade. This cascade leads to the secretion of chemokines and cytokines, notably type 1 interferons, a group which includes IFNα and IFNβ. These signaling molecules attract and activate immune cells within the nearby area.

RIG-I-like receptors (RLRs) are a class of cytoplasmic PRRs. RLRs including retinoic acid-induced gene (RIG-I), melanoma differentiation-associated gene 5 (MDA5), and laboratory of genetics and physiology 2 (LGP2). These molecules share a common structure consisting of a

DExD/H-box helicase domain and a c-terminal repressor domain. Activation of MDA5 or RIG-I leads to the activation of the mitochondrial activator of virus signaling (MAVS) through interaction of caspase activation and recruitment domains (CARDs). The knockout of MAVS protein in experimental mice serves as a useful control for confirming that in absence of this signaling pathway, secretions of IFNα and IFNβ are undetectable and establishing the background signal which can be expected from an assay. Finally, the CARDs' interaction leads to a signaling cascade which results in the activation of two transcription factors, interferon regulatory factor (IRF) 3 and NF-κB. These two transcription factors move from the cytosol to the nucleus and induce transcription of immune response genes, including our secreted cytokines of interest IFN-α and IFN-β.

Knockout of the RLR signaling pathway showed increased mortality compared to wild type (C57BL/6) mice following WNV infection with a systems biology approach revealing hundreds of potential genes involved in host antiviral response[32]. Given the complex regulation of the innate immune system by both host and virus[33], additional pathways and candidate genes seem likely to emerge. While screens have been used previously to investigate this regulation[32,34], the systems biology approach has relied heavily on knock-out and knock-down mouse models and has failed to capture naturally occurring variation which may be more nuanced and provide more relevant insights into immune response.

2.3 Collaborative Cross (CC) Mice

Significant variation in disease processes are known to occur and are of interest to the scientific and medical communities. While it is currently unclear what factors contribute to variable host responses, isolating genetic variation from other factors, such as pathogen dose,

demography, and environment, allows us to focus on the role of genetics in viral disease. Directly studying human immune response to viral infection involves numerous difficulties such as bias against asymptomatic infections, environmental confounders, and cost. However, mouse models offer the opportunity to isolate genetics from other environmental factors. Humans share 99% of mouse genes. Further, mice have well studied phenotypes mirroring a variety of human disease.

Purposed in 2004 by Churchill et al., Collaborative Cross (CC) refers to a panel of recombinant inbred mouse strains which can be used to facilitate the analysis of complex traits and parallel the diversity of wild populations while keeping the ability to replicate experiments with genetically "identical" individuals over time[35]. Eight founder strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, NZO/HlLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ) were bred with the intention of providing 1000 CC strains capturing 90% of phenotypic variation within the mouse population[35]. Currently there are 150 CC strains in existence, though some of these strains are still within the inbreeding process (current generation is less than F_{20})[36]. This limited number of mice reduced resolution and power and does not yet have the ability to detect non-additive effects such as dominance or epistasis. However, research utilizing Collaborative Cross strains for QTL mapping or related genome wide association studies has already been done to study the association between phenotypic variation and genetic loci[37–41], including susceptibility to viral disease[3,42,43].

Additional validations have been done on these CC strains to confirm variation relevant to various areas of study [35,44]. Specifically, diversity in lymphocytes and antigen-presenting cells has been characterized in pre-CC mice (F_5 mice) and used for QTL analysis[45]. Notably, CD11c+ /

CD11b – dendritic cell MFI varied greatly between tested CC strains suggesting that immune response may vary due to immunophenotypic population characteristics rather than variation in signaling pathways within the cell themselves.

The Collaborative Cross model has been shown to mirror West Nile virus disease outcomes in humans[46]. Graham et al. measured disease states in F_1 CC mice following WNV infection including weight loss, clinical score, viral load, IFN-β secretion, and T-cell response. A wide range of susceptibilities were observed including asymptomatic, symptomatic, and symptomatic with center nervous system involvement[46]. Further, this study found that CC strain genetic variation in *Oas1b*, a gene with a well-established association to WNV susceptibility[47], was associated with WNV disease state. CC models should be considered validated for application in the study of immune systems and viral disease.

2.4 Quantitative Trait Locus (QTL) Analysis

Quantitative trait locus (QTL) analysis is an increasingly common statistical method of taking a continuous phenotypic outcome and associating it with a high dimensional set of genetic loci given a group of genetically diverse individuals[48]. While these methods have been utilized often in agricultural studies, the method's success has been mixed[49]. Current limitations on this method include assumption of single allele additive genetic traits, very large sample size requirements, and poor resolution[50]. Multiallelic, dominant traits as well as additive effects are likely to be more common, especially when looking at complex traits such as immune response[51]. Current methods of QTL incorporate environmental and epistatic interaction effects but have limited capacity to control for interaction of multiple alleles for the same trait[52].

Collaborative cross studies have previously made use of QTL analysis [37,38,38,45]. However,

most of these continue to be underpowered due to limited resources and strain number[53,54]. A R processing pipeline has been developed specifically for CC mice and their counterparts, diversity outbred mice. The DOQTL processing pipeline utilizes a linear mixed model with sex and batch covariates and can be used to fit an additive SNP model[55]. Genotypes from founder strains can be estiamted for each of the CC lines using probabilistic imputation[56]. This pipeline has limitations, mainly the absence of built-in checks for normality, poor variance estimation, and analysis based solely on a single phenotype. This single phenotypic input has been previously noted to be problematic for understanding effects which may be regulated at by transcriptional, translational, and degradation processes[57]. Further, complicating the analysis is the need to understand the causal relationship which may involve the interaction of multiple pathways on multiple levels within the cell to interpret gene interactions.

Specialized QTL analysis methods have been proposed for viral studies, which tend to be more dynamic systems than that observed in traditional QTL studies[58,59]. Additional procedures, using the SCOPA software, can make use of multiple correlated phenotypes, especially useful in measuring cytokine responses which tend to be highly correlated[60]. The limited capabilities of current CC QTL analyses means that results from these techniques, especially in underpowered scenarios, should be taken as suggestive of relationships to loci rather than indicative of any novel regulators. In this same sense, a large number of candidate genes will be missed in these approaches[53]. While this is still an emerging field, the potential of these techniques for understanding simple to complex phenotypic traits holds great promise and should be explored in public health genetic research.

2.5 Multiplex Bead Assays

Fluorescence emission is a common assay strategy within the life sciences and can be seen a variety of techniques including immunohistochemistry, flow cytometry, and microarrays. Recently, fluorescence emission has been applied within multiplex bead assays. The most common commercial system is Luminex. While glass and polymer beads have been used since the late 1970s for antigen detection, a reliable system for identifying between beads in a mixed population was only recently established using polymer beads coated with a known ratio of two fluorescent dyes with antigen detection on a third fluorescent channel[61]. Over one hundred unique bead types with different detection antibodies attached have been used simultaneously to detect antigens of interest within the same sample[62]. This assay has also been rigorously tested and proven comparable to the previous gold standard for antigen and antibody detection, the ELISA[63].

Multiplex bead assay results are typically reported as median fluorescent intensity which are then normalized based off the median fluorescent intensities from a known standard curve and blank samples. General guidelines recommend collecting at least 30 beads per bead type per well to obtain this median and ideally closer to one hundred beads per bead type per well. While the median fluorescent intensity has been used consistently and to good effect[62,64], differences in variation between bead batches, sample preparation strategies, and replicate wells have been noticed[62,63,65]. Some researchers have proposed using individual bead level data to measure confidence around each reported value[66] or utilizing a trimmed subset population of beads[64]. Carry-over between wells within the assay has also been a concern with potential biases introduced, especially with sample replicates run with only horizontally on a

plate[67]. While not directly implemented within the work, mixed effect models in conjunction with normalization based off individual beads within a well has the potential to significantly reduce technical variability stemming from multiplex bead assays and improve the accuracy of sensitive screening applications such as QTL analysis.

METHODS

3.1 Ethics Statement

The mouse studies in this work were performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. All mouse studies were performed at the University of North Carolina (Animal Welfare Assurance # A3410-01) using a protocol approved by the UNC Institutional Animal Care and Use Committee (IACUC). All studies were performed in the manner described and minimized pain and suffering.

3.2 Harvesting of Bone Marrow Derived Dendritic Cells

8-10 week old male Collaborative Cross mice were sacrificed. Hindlimbs were collected at UNC-Chapel Hill and shipped overnight on ice to Emory University for processing the next day. Hind limbs were processed similarly to methods escribed previously[68] using chilled RPMI with 10% heat-inactivated FBS, sodium pyruvate, non-essential amino acids, L-glutamine, and antibiotic/antimitotic (here after referred to as cRPMI) instead of cold PBS for flushing procedures. Briefly, two hind limbs were collected from each euthanized animal ensuring that tibia and femur remained intact by gently dislocating femur from hip joint. Hind limbs were then placed in cRPMI, stored on ice, and shipped overnight for processing. Hind limbs were kept at 4C upon receipt and for the entire processing period. Limbs were briefly washed by dunking in a 70% EtOH solution followed by sterile 1x PBS. Muscle tissue was then removed from the limb leaving bone with visible marrow. The end of the tibia and femur were removed using a clean razor blade. Bones were flushed with chilled cRPMI until clear. Resulting extracted marrow was strained through a 70µm cell strainer. Cell counts were taken with 1% Trypan.

Collected bone marrow cells were then seeded at a concentration of $2x10^6$ cells per 6-well plate with cRPMI plus 20ng/mL GM-CSF. Cells were allowed to rest overnight. Media was changed on days 2, 4, 6 post-harvest. On day 7 post-harvest, cells were collected and reseeded at a concentration of $1x10^5$ per well of a 96 well plate for RIG-I agonist treatment or viral infection. Samples from batch were processed in 3 groups (a,b,c) containing 15, 27, and 27 mice including one BL6 control in addition to MAVS KO controls. A subset of hypo and hyper responders from this batch was then duplicated in batch 2, a single batch of 35 mice with biological replicates with technical quadruplicates for each biological replicate (grouped a-d, e-h).

3.3 RIG-I Agonist Treatment

On day 7 post-harvest, 10ng of RIG-I antagonist derived from the 3'-UTR of hepatitis C virus[69] was transfected per $1x10^5$ cells (per well) of mature BMDCs using an mRNA transfection kit. Cells were incubated for an additional 24 hours. Supernatants were harvested and frozen at -80C. Cell pellets were saved in Buffer RLT (QIAgen) and β-Mercaptoethanol at a final concentration of 0.143M (Mirus).

3.4 WNV and ZikV Infection

On day 7 post-harvest, cells were infected with WNV or ZikV at an MOI of 10 (Vero cell titers) for 1 hour. After 1 hour incubation, inoculum was removed and fresh cRPMI were added to the cells. Cells were incubated for an additional 24 hours. Supernatants were harvested and frozen at -80C. Cell pellets were saved in Buffer RLT (QIAgen) and β-Mercaptoethanol at a final concentration of 0.143M.

3.5 Flow Cytometry

The following mouse anti-human antibodies were purchased from BioLegend or Becton

Dickinson: CD11c (B-Ly6), CD80 (2D10), and CD40 (5C3). Following 10min of Fc receptor blockade on ice (Human TruStain FcX, BioLegend), $1x10^5$ cells were sequentially stained for surface markers and viability (Ghost Dye Red 780, Tonbo Biosciences) for 20min on ice. Cells for surface staining were suspended in 10% FCS/PBS and incubated with antibodies for 20min at 4°C. Cells were washed, fixed with BD Fix buffer, and acquired on a BD LSR II with all analysis performed using FlowJo version 10.

3.6 Focus Forming Assay

Supernatants collected from BMDCs infected with WNV or ZikV were serially diluted 1:10 with DMEM supplemented with 1% FBS and used to infect Vero cells for 1 hour at 37C. Cells and inoculum were overlaid with methylcellulose solution (OptiMEM [Corning], 1% Antibiotic/Antimycotic [Corning], 2% FBS, and 2% methylcellulose [Sigma Aldrich]) and incubated for either 24hr (WNV) or 72hr at 37^oC (ZikV). Following incubation, wells were washed with 1x PBS to remove methylcellulose and fixed with either 2% PFA (WNV) or 1:1 methanol : acetone mixture (ZikV) for 30 minutes at room temperature. Cells were then blocked with 5% milk in PBS for 1hr at RT. Cells were incubated with primary antibody (mouse 4G2 monoclonal antibody) at 1µg/mL in 5% milk in PBS for 2hr at RT. Foci were developed with TrueBlue Peroxidase Substrate (KPL). Plates were read on a CTL-ImmunoSpot S6 Micro Analyzer.

3.7 Multiplex Bead Array

Cytokine analysis was performed on supernatants obtained from $1x10^5$ BMDCs following the indicated treatment conditions using a custom magnetic 2-plex panel with mouse IFNβ and IFNα (eBioscience) per the manufacturer's instructions, and read on a Luminex 100 Analyzer.

Supernatant samples from WNV or ZikV infections were UV inactivated for 30 minutes before cytokine analysis.

3.8 Statistical Analysis

All data manipulations, computations, and visualizations were conducted in RStudio (Version 1.0), R (Version 3.2), and BioConductor (3.4). Transformation procedures included none, $\log_2$ transformation, asinh transformation, Box-Cox transformation, and weighted Box-Cox transformation. Estimations of λ were obtained using the boxcox function in MASS package (7.3). Unweighted Box-Cox transformations were applied using the BoxCox function in forecast package (8.0.) Weighted Box-Cox transformations were applied using the bct function in TeachingDemos package(2.10). Log_2 and asinh functions from base R were used. Normalization methods included loess normalization, quantile normalization, and robust-spline normalization. All normalization strategies use the lumiN function from the lumi package (2.26). Additional manipulations included functions from the following packages: MethComp (1.22), matrixStats (0.51), doBy(4.5), reshape2 (1.4), car (2.4), plyr (1.8), timeDate (3012.100). Bland-Altman plots were generated using the bland.altman.plot function from the BlandAltmanLeh package (0.3). QTL and related analyses were performed using the DOQTL package.

3.9 Evaluation of Normalization and Transformation Methods

Six criteria were used to evaluate the effects of preprocessing, three involving visual inspections of normalization plots and three involving empirical criteria. For visual inspections, graphs were rated on a three degree scale; 1 – poor, 2 – fair, 3 – good. Graphs were blinded and rated by two separate raters. 10 of the 40 plots within each plot category were repeated twice to gain a measure of intra-rater reliability. Cohen's kappa for inter-rater agreement was for 0.94, 0.70,

and 0.73 for Bland-Altman plots, MeanSD plots, and QQ-plots respectively. Intra-rater reliability was perfect for both raters. For the 23 out of 120 plots (5 Bland-Altman plots, 10 MeanSD plots, 8 QQ-plots) which had disagreement on the plot rating, a third rater was used as a tie breaker. This third rater rated each questionable plot twice (order of graphs was randomized) and did so blinded to previous ratings. Intra-rater reliability for this third rater was also perfect.

3.9.1. MeanSD plot

Ranked means were plotted against standard deviation to help visualize homoscedasticity (equal variance). Thus, points should make roughly a straight line close to the x-axis. The following instructions were given to raters:

Rank 3:

- scatterplot parallels the x-axis with low variation
- the standard deviation is stable over the mean of signal intensities

Rank 2:

- scatterplot parallels the x-axis with medium variation
- the standard deviation has no clear trend

Rank 1:

- scatterplot does not consistently parallel the x-axis with medium variation
- the standard deviation has a clear trend

3.9.2. Bland-Altman Plot

Bland-Altman plot visualizes the difference between two measurements against their mean. In this case, the different measurements represent different batches. For batch 1, the means of sub-batches 1a and 1b contained no repeated individuals and were combined and compared to sub-batch 1c. No technical replicates were performed within this batch. Duplicate values between the three batches represent biological replicates. For batch 2, two biological replicates were performed with four technical replicates within each biological replicate. Averages of normalized and transformed values for each biological replicate were compared. For batch 1 and 2, average of replicates from batch 1 and average of replicates from batch 2 were compared. The following instructions were given to raters:

Rank 3:

- mean lies at or very close to zero
- variation is constant around the mean

Rank 2:

- mean lies somewhat far from zero
- variation appears to form a trend around the mean

Rank 1:

- mean lies very far from zero
- variation appears to form a trend or funnel around the mean

3.9.3. QQ-Plot

Quantile-quantile plots help diagnose how well measurements align to a theoretical distribution (in this case the normal distribution). The more closely points follow the line, the more closely they follow the expected distribution.

The following instructions were given to raters:

Rank 3:

- Points closely follow line
- Line bisects graph
- Points do not have clear “steps”

Rank 2:

- Points someone closely follow the line but may have some small departures at more extreme values
- Line roughly bisects graph
- Points do not have clear “steps”

Rank 1:

- Points do not follow line
- Line does not dissect graph at all
- Clear steps are visible

3.9.4. Skewness

Skew was calculated as $\mathrm{S} = \left| \log\left(\frac{x_{0.975} - x_{0.5}}{x_{0.5} - x_{0.025}} \right) \right|$. Using this estimation of skewness, a perfectly symmetric distribution would have an S-value equal to zero. For this ranking, an S-value less than 0.5 was considered acceptable.

3.9.5. Tail Length

Tail length was calculated as $\mathrm{T} = \left(\frac{x_{0.975} - x_{0.025}}{x_{0.875} - x_{0.123}} \right)$. Using this estimation of tail length, a normal distribution would have a T-value equal to 1.704. For this ranking, a T-value between 1.5 and 2.5 was considered acceptable.

3.9.6. Kurtosis

Kurtosis is a measure of how heavy or light tailed a distribution is. Kurtosis was calculated using the timeDate function which utilized the Pearson coefficient (K) of kurtosis. Mesokurtic (normal) curves have K = 3. K > 3 is more peaked than normal resulting in a leptokurtic curve. K < 3 is less peaked resulting in a platykurtic curve. Values greater than zero and less than 6 were considered preferable.

3.10 QTL Analysis

All QTL analyses were performed with the DOQTL package in R[70]. Compiled genomes for CC strains were kindly provided by Dr. Martin Ferris at UNC-Chapel Hill. The DOQTL package makes use of Haley-Knott regression. Briefly, a linear mixed model is fitted on mouse diplotype, the

specific combinations of SNPs occurring on a pair of chromosomes including sex as a covariate and a random effect term is included to correct for polygenic covariance. This is considered the full model. A null model without the diplotype term is compared to the full model resulting in a likelihood ratio for the points between marker loci (SNPs). LOD score ($-\log_{10}$(p-value)) is calculated for each loci, and loci sequence is approximated from founder strain sequence associated with marker loci probabilities. This regression method assumes equal variance and normal distribution across phenotypes[63].

3.11 LOD Threshold

LOD thresholds were based off 1000 permutations sampling sixty-one phenotypes from a Gaussian distribution from all provided genomes. Threshold was set at the point where the proportion of permutations in which a loci was correctly detected on any autosome was less than or equal to 0.10. LOD threshold at alpha of 0.90 for both cytokines was 6.63.

RESULTS

4.1 Transformation and Normalization

Large batch effects can be seen when looking at raw values for both cytokines (Figure 4.1). Raw values from IFNα and IFNβ also have significant differences in magnitude of variance observed (Table 4.1). Particularly notable transformations are the log2 and asinh which resulted in significant improvement in reducing skew of distribution, especially in the case of IFNα (Table 4.2). Both unweighted and weighted Box-Cox methods increased skew and shifted tail length and kurtosis away from true normal relative to performing no transformations. Quantile normalization proved most successful of the normalization strategies applied, especially when applied in conjunction with log2 transformation (Tables 4.2, 4.3). Loess normalization also performed well in conjunction with log2 and asinh transformations (Tables 4.2, 4.3). Loess and reduced spline normalization proved inappropriate for use when the majority of replicate data was missing as was the case for batch 1 and batch 1 and 2 combined (data not shown). Log2 and asinh transformation worsened tail length and kurtosis, as did both quantile and loess normalization. Qualitative and quantitative measures largely mirrored on another. QQ-Plots showed the greatest departure from other measures, in particular receiving mid ratings to high ratings for log2 and asinh transformations within Batch 1 across both IFNα and IFNβ. Log2-quantile strategy was determined to be the most promising based off of observed performance for both cytokines. Visualizing normalized data using this selected strategy, one can see overall decreased variability within a batch effect (Figure 4.1).

4.2 QTL Analysis

QTL analysis proved highly sensitive to the log2-quantile strategy employed. Comparing log2-quantile manipulated data with raw data, we see major changes in both LOD values and the loci of interest themselves. While we see most of these locations are lost following normalization, clear peaks of LOD are seen for both cytokines (Figure 4.2). Using an LOD loci cutoff of 6.63, no loci were statistically significant. However, clear peaks were seen within the QTL visualization. The peaks with the greatest LOD values were analyzed further and the highest peaks for both cytokines contained potential genes of interest (Table 4.1 and Figure 4.2). Potential genes and transcriptional modulators contained within loci of interest were not investigated further. Unfortunately, no founder strains were clearly associated with the loci of interest (Figure 4.2).

4.3 Genes of Interest – IFNα

The loci of interest with the greatest LOD resulting from QTL analysis of normalized IFNα cytokine phenotypes was located on chromosome 12 between base pairs 103310975 and 103406350. This region contained three genes: Fam181a, Asb2, and Otub2. Fam181a is a poorly characterized gene which has been implicated in nervous system development and function[73]. Asb2 promotes protein ubiquitination by forming non-canonical E3 ligase complexes[74–76]. E3 ligase complexes have been previously implicated in viral infection with viral genes and host genes manipulated by viruses resulting to both enhanced and diminished ubiquitination which benefit the viral infections[77]. Most promisingly, Otub2 has been shown to inhibit virus induced cytokine signaling with Sendai virus infection in mice[78]. In addition, Otub2 has been found to

impact ubiquitination on the cellular level resulting in preferential changes in DNA repair mechanism[79].

4.4 Genes of Interest – IFNβ

The loci of interest with the greatest LOD resulting from QTL analysis of normalized IFNβ cytokine phenotypes was located on chromosome 4 between base pairs 65124174 and 65616238. This region also contained three genes: Pappa, Astn2, and Trim32. Pappa has been linked to NFκB responses within various cancers[80–82]. The NFκB inflammatory signaling response is engaged by the RLR signaling pathway as well. Astn2 has been studied primarily in relation to neural malformations[83–85]. Somewhat strangely, mutations in this gene have been shown to alter hair follicle orientation[86]. Trim32 has been implicated within RLR signaling pathways within fish in response to iridovirus and nodavirus[87]. In addition, it has been shown to regulated influenza A virus via ubiquitination within mice[88].

DISCUSSION

5.1 Transformation and Normalization

The poor performance of Box-Cox transformation was surprising given the previous reported success with this method[56]. This poor performance may be due to the small number of replicates. Only quantile normalization was capable of handling the unbalanced sampling present within batch 1 and between batches 1 and 2. The aims of these methods were two fold, to create a more normalized distribution to fit the assumptions of the statistical model and to attempt to reduce the batch effects seen within the data. Bland-Altman plots in particular proved a useful tool in enabling both the visualization of variance and the difference remaining between batches following normalization. Further, it is worth noting that different normalization methods had different levels of success with different antigens tested. While quantile normalization consistently offered improved normality measures compared to raw data, like most of its fellow methods, this improvement appeared to come at the cost of tail length and kurtosis. A more detailed analysis of variation among these elements within regression would be necessary before a proper weight could be assigned.

The grading of the qualitative measures was optimized for reliability rather than precision. The lack of categorization for quantitative measures was purposeful as it allowed for more nuanced analysis. The final selection of transformed and normalized data for testing relative to raw data in QTL analysis was however largely subjective and based off of the author's assessment of all six measures rather than a quantitative ranking.

Normalization strategies were based largely off of previous techniques developed for microarray data. Some key differences hinder their implementation in the context of MBA, primarily smaller number of analytes and replicates. Other issues with applying normalization

strategies in this study included the unbalanced sampling with a non-randomly selected number of individuals receiving biological replicates and/or technical replicates.

5.2 QTL Analysis

One of the strengths of this methodology was using permutations to derive the LOD cutoffs. This method is more reflective of statistical significance than setting an a priori, arbitrary cutoffs independent of phenotypes utilized[64]. This method did not provide a clear cutoff for founder strain effects, which the author did decide a prior to set as LOD cutoff minus one. Due to the use of SNPs to estimate the probability of sequence inheritance from founder strains, it is unclear whether gene sequences are conserved between founders and thereby the LOD of founder effects are likely to be an underestimation of allele affects.

5.3 Technical Variation

The major weakness of this study was the variance observed between batches, biological replicates, and technical replicates. The extreme level of this variation can be seen when ranking random subsets of combined data from batches 1 and 2. Here different random samples of phenotypic measures taken from the data set will lead to strains being ranked alternatively within the lowest or highest quartiles of the observed phenotypes. One solution may be to include additional controls beyond the single BL6 within each batch and then normalize to each batch using one of various statistical techniques including multivariate regression or Bayesian approaches. Another solution would be to more thoroughly incorporate random biological replicates from various CC lines within different batches. This would have the added advantage of providing a greater number of biological and replicates for phenotypes of interest.

While there is likely to have been numerous sources of variation combining and compounding one another to result in the observed variance, additional pilot experiments suggest that further purification of the bone marrow extracted cells will greatly reduce technical variation within a biological replicates within batches (data not shown). This suggests the bone marrow derived dendritic cells used within this work were actually a highly mixed population of cells. Therefore, each technical replicate had a unique level of cytokine production dependent on the distribution of cells within the well population and how those cells interacted to elicit a response to RIG-I antagonist or viral infection. Additional flow cytometry experiments following BMDC culturing could determine if this is the case.

5.6 Future Directions and Concluding Remarks

Statistical analysis of MBAs is still in its early stages. While QTL analysis has been frequently utilized in agricultural research, this method faces serious power limitations when conducting screening research in current CC strains. Therefore, it is not surprising that our screening yielded no statistically significant results at an alpha of 0.90. Yet, we found the clear formation of peaks graphing LOD across the murine genome suggestive that results obtained may still be useful moving forward and that significance may be a poor tool for gauging screening outcomes with phenotypic expression with large variation. Both of the highest LOD peaks identified within the QTL analysis yielded potential genes of interest previously associated with innate immune response to viral infection. While both cytokines should be closely correlated, the loci mapped varied greatly. Both loci suggested that rather the RLR signaling pathway, ubiquitination following viral infection may be more greatly influenced by host genetics. The combined use of these methods for understanding how genetic variation

within a population may contribute to host immune response and viral susceptibility is novel and requires a carefully considered statistical approach. Future work includes looking at expression of identified genes of interest at the transcriptional and translational level, trying to reproduce these results *in vivo*, and repeating analysis including founder strains to see if a specific allele resulting in phenotypic change can be identified.

Treatment	Assay	Batch	Mean (Prop)	SD	Min	Max
Untreated	CD86	1	0.364	0.772	-1	3.9
		1a	0.759	0.916	-0.3	3.9
		1b	0.148	0.576	-0.9	1.6
		1c	0.040	0.470	-1	1
		2	0.051	0.555	-1.3	1.2
		1,2	0.259	0.719	-1.3	3.9
	CD40	1	0.183	0.847	-3.9	2.1
		1a	0.500	0.599	-3.9	2
		1b	-0.011	1.096	-0.3	1.7
		1c	-0.040	0.521	-3.9	2
		2	0.743	0.795	-0.8	0.7
		1,2	0.371	0.868	-1.1	2.1
	IFNa	1	14.043	1.344	11.5	17
		1a	13.926	1.452	11.5	17
		1b	14.222	1.273	12	16.5
		1c	13.933	1.321	12	16
		2	14.861	5.205	8	32
		1,2	14.596	4.359	8	32
	IFNb	1	10.043	0.808	8	12
		1a	9.963	0.678	9	11
		1b	10.130	0.967	8	12
		1c	10.033	0.743	9	11
		2	44.528	144.731	16.5	1510.5
		1,2	33.357	119.963	8	1510.5
RIG-I	IFNa	1	2130.775	2948.452	15	14261.5
		1a	3803.926	3633.210	264	14261.5
		1b	1498.833	2010.737	88	6951
		1c	256.600	518.827	31	2074.5
		2	2610.583	4892.790	16	20240.5
		1,2	2455.153	4357.395	16	20240.5
	IFNb	1	3300.848	2928.603	9	11155
		1a	5329.148	2780.489	1796.5	11155
		1b	2510.593	2444.532	118	9860
		1c	1072.367	1320.358	119	5051.5
		2	2674.847	4353.354	43	20419
		1,2	2877.636	3952.303	43	20419
WNV	IFNa	1	68.210	278.344	15	2298.5
		1a	103.185	438.750	16	2298.5
		1b	58.722	86.627	17	380
		1c	22.333	8.983	15	43.5
	IFNb	1	32.717	74.498	9	574
		1a	34.648	107.857	10	574
		1b	38.963	51.439	9	265.5
		1c	18.000	11.128	10	52.5
	FFA	1	290084.058	347640.846	13600	1440000
		1a	500740.741	434898.803	44000	1440000
		1b	175518.519	217716.275	14200	920000
		1c	117120.000	69087.307	13600	224000
ZikV	IFNa	1	20.319	2.501	16	31
		1a	19.778	1.577	16	22.5
		1b	21.148	3.379	17	31
		1c	19.800	1.545	18	23
	IFNb	1	12.210	2.832	10	28
		1a	11.370	0.674	10	13
		1b	13.241	4.246	10	28
		1c	11.867	1.043	10	14
	FFA (Proportion)	1	(0.580)			
		1a	(0.629)			
		1b	(0.440)			
		1c	(0.530)			

Table 4.1. Summary Characteristics of Phenotypic Variables. CD86 and CD40 values are the mean of log2 transformed median fluorescent intensities captured in CD11c+ population of bone marrow isolated cells measured using flow cytometry. Values were normalized to BL6 controls. IFNα and IFNβ are the mean of median fluorescent intensities of cytokines measured by multiplex bead assay with background subtracted. Focus forming assay (FFA) values are the mean of the focus forming units per milliliter. FFA performed for ZikV viruses are reported as a proportion of assays with foci present.

Batch(es)	Transformation	Normalization	Skew	Tail Length	Kurtosis	Bland-Altman	MeanSD	QQ-Plot
1	None	None	3.140	1.930	4.510	1	1	1
		Quantile	2.820	1.910	2.910	1	1	1
	Log2	None	0.600	1.430	-1.830	1	1	2
		Quantile	0.520	1.420	-1.890	1	1	2
	asinh	None	0.600	1.430	-1.870	1	1	2
		Quantile	0.520	1.420	-1.920	1	1	2
	Box-Cox	None	7.500	4.500	-2.070	1	1	1
		Quantile	2.600	1.600	-2.070	1	1	2
	Weighted Box-Cox	None	7.500	4.500	-2.070	1	1	1
		Quantile	2.570	1.740	-2.570	1	1	1
2	None	None	5.230	1.740	6.340	1	1	1
		Quantile	4.920	1.750	6.760	1	1	1
		Loess	5.070	1.730	6.400	1	1	1
		RSN	6.370	1.780	6.050	1	1	1
	Log2	None	0.470	1.280	-1.720	2	2	2
		Quantile	0.430	1.250	-1.730	3	2	2
		Loess	0.400	1.300	-1.700	2	2	2
		RSN	0.730	1.220	-1.540	3	2	2
	asinh	None	0.480	1.280	-1.820	2	2	2
		Quantile	0.430	1.250	-1.820	3	2	2
		Loess	0.430	1.290	-1.810	2	2	2
		RSN	1.390	1.150	-1.800	2	2	2
	Box-Cox	None	7.180	8.350	-2.260	1	1	1
		Quantile	7.110	3.310	-2.270	1	1	1
		Loess	7.320	9.230	-2.260	1	1	1
		RSN	7.310	2.020	-2.260	1	1	1
	Weighted Box-Cox	None	7.180	8.350	-2.260	1	1	1
		Quantile	7.110	3.310	-2.270	1	1	1
		Loess	7.320	9.230	-2.260	1	1	1
		RSN	6.920	1.990	-2.260	1	1	1
1,2	None	None	3.970	2.080	6.890	1	1	1
		Quantile	4.370	2.580	7.530	1	1	1
	Log2	None	0.001	1.408	-1.804	2	2	2
		Quantile	0.287	1.561	-1.744	2	2	3
	asinh	None	0.001	1.408	-1.849	2	2	2
		Quantile	0.290	1.560	-1.805	2	2	3
	Box-Cox	None	7.440	9.780	-2.070	1	1	1
		Quantile	6.640	6.460	-2.070	1	1	2
	Weighted Box-Cox	None	7.440	9.780	-2.070	1	1	1
		Quantile	1.940	1.930	-2.670	1	1	1

Table 4.2. Performance of Pre-Processing Strategies in IFNα. All methods were reported as described above. Highlighted cells are those considered to have promising performance as normalization approaches.

Batch(es)	Transformation	Normalization	Skew	Tail Length	Kurtosis	Bland-Altman	MeanSD	QQ-Plot
1	None	None	1.720	1.458	0.323	1	1	1
		Quantile	1.892	1.825	0.179	1	1	1
	Log2	None	1.034	1.486	-1.979	1	1	2
		Quantile	0.175	1.691	-2.006	1	1	3
	asinh	None	1.034	1.486	-1.992	1	1	2
		Quantile	0.175	1.691	-2.016	1	1	3
	Box-Cox	None	6.263	3.669	-2.070	1	1	1
		Quantile	4.562	3.810	-2.070	1	1	1
	Weighted Box-Cox	None	6.263	3.669	-2.070	1	1	1
		Quantile	1.588	1.215	-6.000	1	1	1
2	None	None	4.522	1.957	6.827	1	2	1
		Quantile	4.436	2.036	5.943	1	2	1
		Loess	4.450	1.989	6.316	1	2	1
		RSN	4.926	1.882	6.673	1	3	1
	Log2	None	0.798	1.262	-1.987	2	1	2
		Quantile	0.764	1.273	-1.988	3	1	2
		Loess	0.779	1.274	-1.985	2	1	2
		RSN	0.877	1.257	-1.883	3	1	2
	asinh	None	0.798	1.262	-2.036	2	2	2
		Quantile	0.764	1.273	-2.036	3	2	2
		Loess	0.781	1.272	-2.035	2	2	2
		RSN	1.210	1.219	-1.975	2	2	2
	Box-Cox	None	3.742	1.938	-2.267	1	1	1
		Quantile	3.599	1.558	-2.267	1	1	1
		Loess	3.742	1.938	-2.267	1	1	1
		RSN	4.000	1.820	-2.267	1	1	1
	Weighted Box-Cox	None	3.742	1.938	-2.267	1	1	1
		Quantile	3.599	1.558	-2.267	1	1	1
		Loess	3.742	1.938	-2.267	1	1	1
		RSN	3.952	1.759	-2.267	1	1	1
1,2	None	None	2.272	1.578	2.098	1	1	1
		Quantile	3.651	2.061	2.585	1	1	1
	Log2	None	0.581	1.392	-1.956	2	2	2
		Quantile	0.547	1.406	-1.922	2	3	2
	asinh	None	0.581	1.392	-1.973	2	1	2
		Quantile	0.547	1.406	-1.946	3	1	2
	Box-Cox	None	5.925	2.254	-2.069	1	1	1
		Quantile	4.003	2.575	-2.069	1	1	1
	Weighted Box-Cox	None	5.925	2.254	-2.069	1	3	1
		Quantile	2.538	1.215	-3.000	1	3	1

Table 4.3. Performance of Pre-Processing Strategies in IFNβ. All methods were reported as described above. Highlighted cells are those considered to have promising performance as normalization approaches.

IFNβα						
Chromosome	start	stop	strand	Name	Dbxref	mgiName
12	103310975	103317065	+	Fam181a	NCBI_Gene:100504156	family with sequence similarity 181, 2c member A
12	103321142	103356001	-	Asb2	NCBI_Gene:65256	ankyrin repeat and SOCS box-containing 2
12	103388682	103406350	+	Otub2	NCBI_Gene:68149	OTU domain, 2c ubiquitin aldehyde binding 2
IFNβ						
Chromosome	start	stop	strand	Name	Dbxref	mgiName
4	65124174	65357509	+	Pappa	NCBI_Gene:18491	pregnancy-associated plasma protein A
4	65380803	66404611	-	Astn2	NCBI_Gene:56079	astrotactin 2
4	65604986	65616238	+	Trim32	NCBI_Gene:69807	tripartite motif-containing 32

Table 4.4. QTL Results for Genes of Interest. Genes contained within highest significance loci of interest for log2 transformed, quantile normalized cytokine data. Neither loci meet the 6.63 significance level cut off for LOD (IFNα, LOD = 6.12; IFNβ, LOD = 5.67)

Procedure		Inter-batch variation	Intra-strain variation	Inter-assay variation	Inter-plate variation	Intra-plate variation
Harvesting Bone Marrow		Shipping conditions	Bone integrity			
			Age			
			Sex			
			Uncaptured heterogeneity in inbred lines			
			Health status			
Seeding and Culturing Cells		Media preparations	Processing order			
			Cell counting accuracy			
		Processing time	Health and derivation of cells			
			Pipetting accuracy			
			Cell activation			
Treatment	Infection	Viral titers	Cell population viral susceptibility			
			Pipetting accuracy			
			Cell death			
	RIG-I Agonist	Reagent preparation	Pipetting accuracy			
Supernatant Harvesting		Harvesting time		UV inactivation (viral only)	Harvesting order	Defrosting time
						Temperature
						Volume harvested
MBA				Kit differences	Assay standards	Accuracy of pipetting
				Reagent preparation	Bead concentrations	Spillage or leakage between wells
				Instrument calibration and performance	Instrument temperature	
				Incubation times	Wash effectiveness	Carryover from previous wells
					Processing order	
Potential Control		Pooled, repeated samples	Biological replicates	Pooled, repeated samples	Pooled, repeated samples	Random allocation
		Standards	Sibling matching	Standards	Standards	Bookend standard dilutions
		Dispersed controls				

Table 5.1. Sources of Technical Variation. Potential sources of technical variation for each procedure are listed grouped by sources of variation. The last row in the Procedure column provides potential controls for use in a mixed effect model or which might be better controlled in the future.

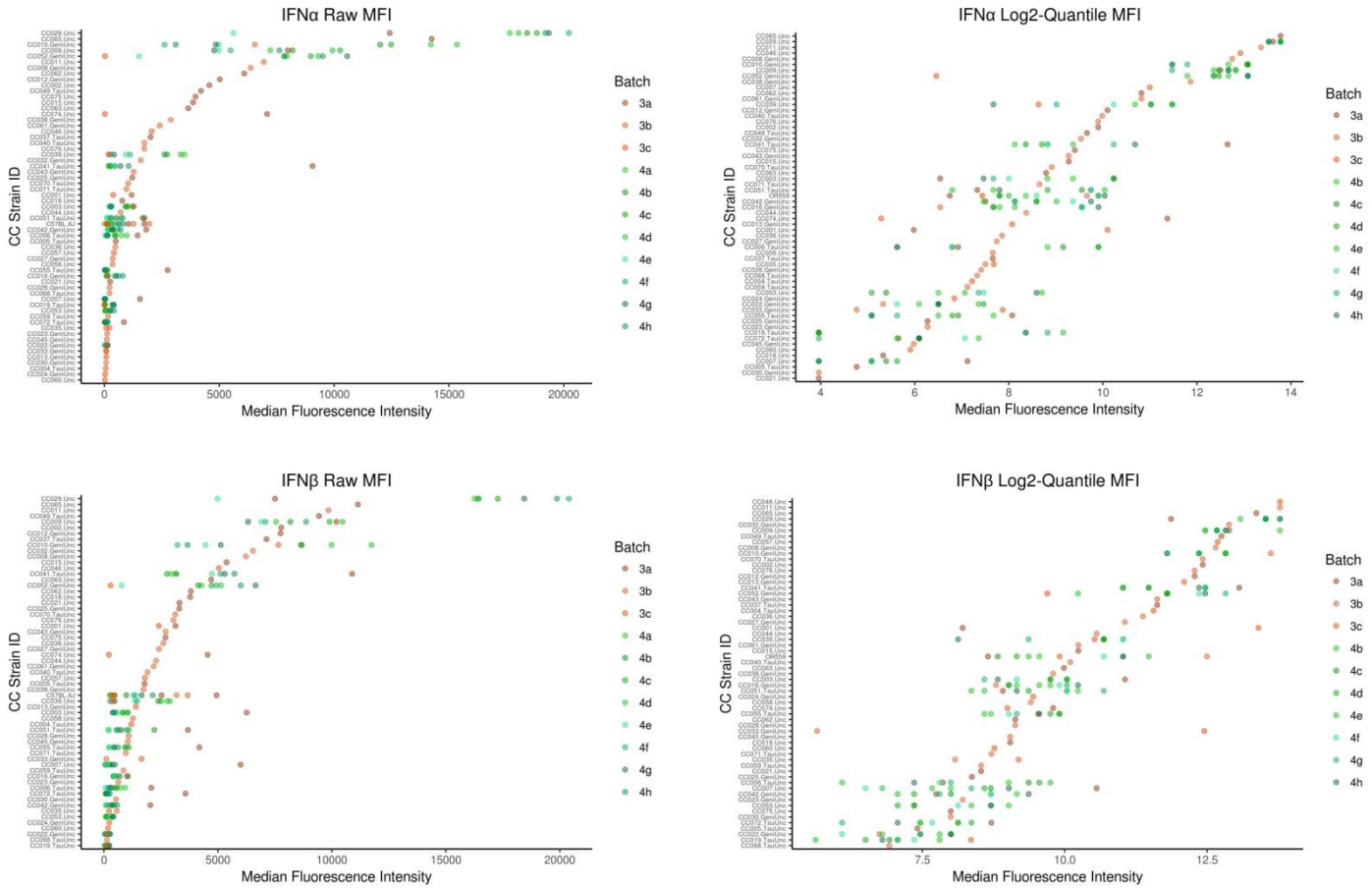

Figure 4.1 Distributions of IFNβ and IFNα by Strain. Replicates of each strain are shown with ordering of strains by row or normalized cytokine response. Median fluorescent intensity reported had the background values subtracted. 2a-d and 2e-h were technical replicates within each group. All other replicates were biological.

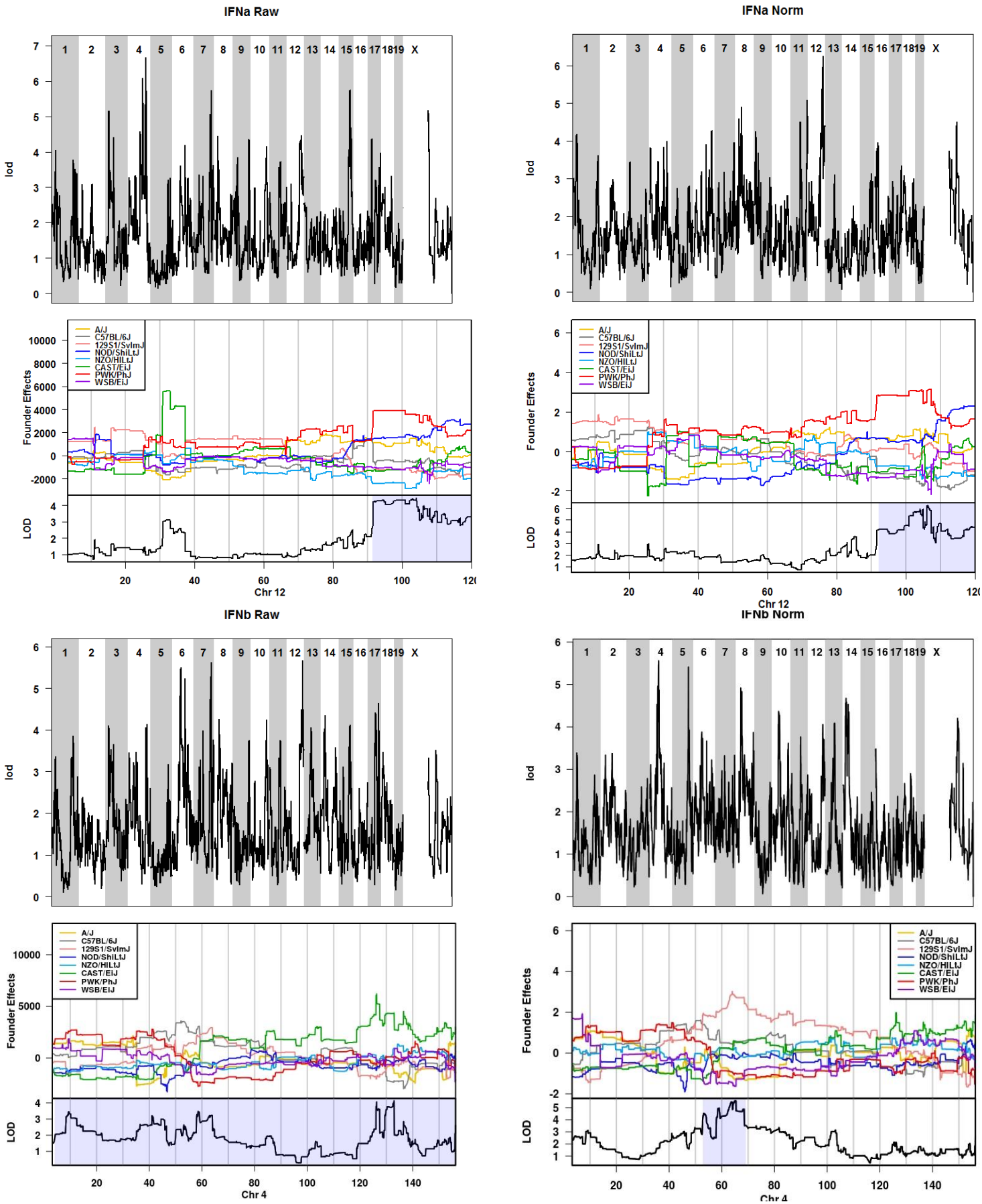

Figure 4.2 QTL Analyses of Selected Method. QTL analyses using both raw and normalized values were conducted. Founder effects were analyzed for both raw and normalized values for chromosomes which contained the loci with greatest LOD value in normalized analysis, Chromosomes 4 and 12 for IFNα and IFNβ respectively.

REFERENCES

1. Stein, C. M. Genetic Epidemiology of Tuberculosis Susceptibility: Impact of Study Design. *PLOS Pathog.* **7,** e1001189 (2011).
2. Doeschl-Wilson, A. B. *et al.* Implications of Host Genetic Variation on the Risk and Prevalence of Infectious Diseases Transmitted Through the Environment. *Genetics* **188,** 683–693 (2011).
3. Rasmussen, A. L. *et al.* Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science* **346,** 987–991 (2014).
4. CDC. *West Nile Virus Disease Cases and Presumptive Viremic Blood Donors by State – United States, 2015*.
5. CDC. *ZikV Case Counts in the US*.
6. Quicke, K. M. *et al.* Zika Virus Infects Human Placental Macrophages. *Cell Host Microbe* **20,** 83–90 (2016).
7. Dowall, S. D. *et al.* A Susceptible Mouse Model for Zika Virus Infection. *PLoS Negl. Trop. Dis.* **10,** e0004658 (2016).
8. Samuel, C. E. Host genetic variability and West Nile virus susceptibility. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11555–11557 (2002).
9. Lessler, J. *et al.* Assessing the global threat from Zika virus. *Science* aaf8160 (2016). doi:10.1126/science.aaf8160
10. Elbahesh, H. & Schughart, K. Genetically diverse CC-founder mouse strains replicate the human influenza gene expression signature. *Sci. Rep.* **6,** 26437 (2016).
11. Westaway, E. G. Flavivirus. *Springer Index Viruses* 306–319 (2002).

12. Jeffrey Root, J. West Nile virus associations in wild mammals: a synthesis. *Arch. Virol.* **158,** 735–752 (2013).
13. Lim, S. M. *et al.* Susceptibility of Carrion Crows to Experimental Infection with Lineage 1 and 2 West Nile Viruses. *Emerg. Infect. Dis.* **21,** 1357–1365 (2015).
14. Pello, S. J. & Olsen, G. H. Emerging and Reemerging Diseases of Avian Wildlife. *Veterinary Clin. North Am. Exot. Anim. Pract.* **16,** 357–381 (2013).
15. Marschang, R. E. Viruses Infecting Reptiles. *Viruses* **3,** 2087–2126 (2011).
16. Weissenböck, H., Hubálek, Z., Bakonyi, T. & Nowotny, N. Review: Zoonotic mosquito-borne flaviviruses: Worldwide presence of agents with proven pathogenicity and potential candidates of future emerging diseases. *Vet. Microbiol.* **140,** 271–280 (2010).
17. Dobler, G. Review: Zoonotic tick-borne flaviviruses. *Vet. Microbiol.* **140,** 221–228 (2010).
18. Grogan, L. F. *et al.* Surveillance for Emerging Biodiversity Diseases of Wildlife. *PLoS Pathog.* **10,** 1–4 (2014).
19. Sejvar, J. J. Clinical Manifestations and Outcomes of West Nile Virus Infection. *Viruses* **6,** 606–623 (2014).
20. ROTENBERRY, J. T. & UNFRIED, T. M. Avian Ecology and Conservation in an Urbanizing World. *Auk Univ. Calif. Press* **119,** 889 (2002).
21. Suthar, M. S., Diamond, M. S. & Gale, M. West Nile virus infection and immunity. *Nat. Rev. Microbiol.* **11,** 115–128 (2013).
22. Fauci, A. S. & Morens, D. M. Zika Virus in the Americas--Yet Another Arbovirus Threat. *N. Engl. J. Med.* **374,** 601–604 (2016).

23. Garcia, E. *et al.* Zika virus infection: global update on epidemiology and potentially associated clinical manifestations. *Wkly. Epidemiol. Rec.* **91,** 73–81 (2016).
24. Fellner, C. Zika Virus: Anatomy of a Global Health Crisis. *Pharm. Ther.* **41,** 242–253 (2016).
25. Jaenisch, T. *et al.* Risk of microcephaly after Zika virus infection in Brazil, 2015 to 2016. *Bull. World Health Organ.* **95,** 191–198 (2017).
26. Cuevas, E. L. *et al.* Preliminary Report of Microcephaly Potentially Associated with Zika Virus Infection During Pregnancy - Colombia, January-November 2016. *MMWR Morb. Mortal. Wkly. Rep.* **65,** 1409–1413 (2016).
27. LaBeaud, A. D., Bashir, F. & King, C. H. Measuring the burden of arboviral diseases: the spectrum of morbidity and mortality from four prevalent infections. *Popul. Health Metr.* **9,** 1 (2011).
28. John, C. C. *et al.* Global research priorities for infections that affect the nervous system. *Nature* **527,** S178–S186 (2015).
29. Ishikawa, T., Yamanaka, A. & Konishi, E. A review of successful flavivirus vaccines and the problems with those flaviviruses for which vaccines are not yet available. *Vaccine* **32,** 1326–1337 (2014).
30. How NIAID Is Addressing Zika Virus | NIH: National Institute of Allergy and Infectious Diseases. Available at: https://www.niaid.nih.gov/diseases-conditions/addressing-zika. (Accessed: 8th April 2017)
31. Barouch, D. H., Thomas, S. J. & Michael, N. L. Prospects for a Zika Virus Vaccine. *Immunity* **46,** 176–182 (2017).

32. Suthar, M. S. *et al.* A Systems Biology Approach Reveals that Tissue Tropism to West Nile Virus Is Regulated by Antiviral Genes and Innate Immune Cellular Processes. *PLoS Pathog.* **9,** (2013).

33. Ma, D. Y. & Suthar, M. S. Mechanisms of innate immune evasion in re-emerging RNA viruses. *Curr. Opin. Virol.* **12,** 26–37 (2015).

34. Korth, M. J., Tchitchek, N., Benecke, A. & Katze, M. G. Systems approaches to influenza-virus host interactions and the pathogenesis of highly virulent and pandemic viruses. *Semin. Immunol.* **25,** (2013).

35. Churchill, G. A. *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* **36,** 1133–1137 (2004).

36. Bogue, M. A., Churchill, G. A. & Chesler, E. J. Collaborative Cross and Diversity Outbred data resources in the Mouse Phenome Database. *Mamm. Genome* **26,** 511–520 (2015).

37. Abu-Toamih Atamni, H., Iraqi, F., Ziner, Y., Wolf, L. & Mott, R. Glucose tolerance female-specific QTL mapped in collaborative cross mice. *Mamm. Genome* **28,** 20 (2017).

38. Levy, R., Mott, R. F., Iraqi, F. A. & Gabet, Y. Collaborative cross mice in a genetic association study reveal new candidate genes for bone microarchitecture. *BMC Genomics* **16,** 1–14 (2015).

39. Vered, K., Durrant, C., Mott, R. & Iraqi, F. A. Susceptibility to klebsiella pneumonaie infection in collaborative cross mice is a complex trait controlled by at least three loci acting at different time points. *BMC Genomics* **15,** 865–874 (2014).

40. Atamni, H., Iraqi, F., Botzman, M., Gat-Viks, I. & Mott, R. Mapping liver fat female-dependent quantitative trait loci in collaborative cross mice. *Mamm. Genome* **27,** 565 (2016).
41. Lorè, N. I., Iraqi, F. A. & Bragonzi, A. Host genetic diversity influences the severity of Pseudomonas aeruginosa pneumonia in the Collaborative Cross mice. *BMC Genet.* **16,** 106–106 (2015).
42. Gralinski, L. E. *et al.* Genome Wide Identification of SARS-CoV Susceptibility Loci Using the Collaborative Cross. *PLoS Genet.* **11,** (2015).
43. Ferris, M. T. *et al.* Modeling Host Genetic Regulation of Influenza Pathogenesis in the Collaborative Cross. *PLoS Pathog.* **9,** (2013).
44. Bogue, M. A., Churchill, G. A. & Chesler, E. J. Collaborative Cross and Diversity Outbred data resources in the Mouse Phenome Database. *Mamm. Genome* **26,** 511–520 (2015).
45. Phillippi, J. *et al.* Using the emerging Collaborative Cross to probe the immune system. *Genes Immun.* **15,** 38–46 (2014).
46. Graham, J. B. *et al.* Genetic Diversity in the Collaborative Cross Model Recapitulates Human West Nile Virus Disease Outcomes. *mBio* **6,** (2015).
47. Lucas, M. *et al.* Infection of mouse neurones by West Nile virus is modulated by the interferon-inducible 2'-5' oligoadenylate synthetase 1b protein. *Immunol. Cell Biol.* **81,** 230–236 (2003).
48. Li, H. & Deng, H. Systems genetics, bioinformatics and eQTL mapping. *Genet. Hague* **138,** 915–24 (2010).

49. Bernardo, R. What proportion of declared QTL in plants are false? *Theor. Appl. Genet.* **109,** 419–424 (2004).

50. Roff, D. A. A Centennial Celebration for Quantitative Genetics. *Evolution* 1017 (2007).

51. Mackay, T. F. The genetic architecture of quantitative traits: lessons from Drosophila. *Curr. Opin. Genet. Dev.* **14,** 253–257 (2004).

52. Goffinet, B. & Gerber, S. Quantitative Trait Loci: A Meta-analysis. *Genetics* **155,** 463–473 (2000).

53. van Ooijen, J. W. LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* **83,** 613–624 (1999).

54. Zou, F., Nie, L., Wright, F. A. & Sen, P. K. A robust QTL mapping procedure. *J. Stat. Plan. Inference* **139,** 978 (20090101).

55. Rat Genome Sequencing and Mapping Consortium *et al.* Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat. Genet.* **45,** 767–775 (2013).

56. Bush, W. S. & Moore, J. H. Chapter 11: Genome-wide association studies. *PLoS Comput. Biol.* **8,** e1002822 (2012).

57. Stylianou, I. M. *et al.* Applying gene expression, proteomics and single-nucleotide polymorphism analysis for complex trait gene identification. *Genetics* **178,** 1795–1805 (2008).

58. Luo, J., Hager, W. W. & Wu, R. A differential equation model for functional mapping of a virus-cell dynamic system. *J. Math. Biol.* **61,** 1 (20100101).

59. Sun, L. & Wu, R. Review: Mapping complex traits as a dynamic system. *Phys. Life Rev.* **13,** 155–185 (2015).

60. Mägi, R. *et al.* SCOPA and META-SCOPA: software for the analysis and aggregation of genome-wide association studies of multiple correlated phenotypes. *BMC Bioinformatics* **18,** 1–8 (2017).
61. Tighe, P., Negm, O., Todd, I. & Fairclough, L. Utility, reliability and reproducibility of immunoassay multiplex kits. *Methods San Diego Calif* **61,** 23–29 (2013).
62. Siawaya, J. F. D. *et al.* An Evaluation of Commercial Fluorescent Bead-Based Luminex Cytokine Assays. *PLOS ONE* **3,** e2535 (2008).
63. Eckels, J. *et al.* Quality control, analysis and secure sharing of Luminex® immunoassay data using the open source LabKey Server platform. *BMC Bioinformatics* **14,** 145 (2013).
64. Clarke, D. C., Morris, M. K. & Lauffenburger, D. A. Normalization and Statistical Analysis of Multiplexed Bead-based Immunoassay Data Using Mixed-effects Modeling. *Mol. Cell. Proteomics MCP* **12,** 245–262 (2013).
65. Hanley, B. P. Variance in multiplex suspension array assays: A distribution generation machine for multiplex counts. *Theor. Biol. Med. Model.* **5,** 3 (2008).
66. Jacobson, J. W., Oliver, K. G., Weiss, C. & Kettman, J. Analysis of individual data from bead-based assays ('bead arrays'). *Cytometry A* **69A,** 384–390 (2006).
67. Hanley, B. Variance in multiplex suspension array assays: carryover of microspheres between sample wells. *J. Negat. Results Biomed.* **6,** 6 (2007).
68. Zanoni, I. *et al.* CD14 regulates the dendritic cell life cycle after LPS exposure through NFAT activation. *Nature* **460,** 264–268 (2009).

69. Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J. & Jr, M. G. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* **454,** 523–527 (2008).

70. Gatti, D. M. *et al.* Quantitative Trait Locus Mapping Methods for Diversity Outbred Mice. *G3 Genes Genomes Genet.* **4,** 1623–1633 (2014).

71. Haley, C. S. & Knott, S. A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69,** 315–324 (1992).

72. Rausch, T. K. *et al.* Comparison of pre-processing methods for multiplex bead-based immunoassays. *BMC Genomics* **17,** (2016).

73. Marks, M. *et al.* Analysis of the Fam181 gene family during mouse development reveals distinct strain-specific expression patterns, suggesting a role in nervous system development and function. *Gene* **575,** 438–451 (2016).

74. Heuzé, M. L. *et al.* ASB2 is an Elongin BC-interacting protein that can assemble with Cullin 5 and Rbx1 to reconstitute an E3 ubiquitin ligase complex. *J. Biol. Chem.* **280,** 5468–5474 (2005).

75. Heuzé, M. L. *et al.* ASB2 targets filamins A and B to proteasomal degradation. *Blood* **112,** 5130–5140 (2008).

76. Nie, L. *et al.* Notch-induced Asb2 expression promotes protein ubiquitination by forming non-canonical E3 ligase complexes. *Cell Res.* **21,** 754–769 (2011).

77. Randow, F. & Lehner, P. J. Viral avoidance and exploitation of the ubiquitin system. *Nat. Cell Biol.* **11,** 527–534 (2009).

78. Li, S. *et al.* Regulation of virus-triggered signaling by OTUB1- and OTUB2-mediated deubiquitination of TRAF3 and TRAF6. *J. Biol. Chem.* **285,** 4291–4297 (2010).

79. Kato, K. *et al.* Fine-tuning of DNA damage-dependent ubiquitination by OTUB2 supports the DNA repair pathway choice. *Mol. Cell* **53,** 617–630 (2014).

80. Engelmann, J. C. *et al.* Causal Modeling of Cancer-Stromal Communication Identifies PAPPA as a Novel Stroma-Secreted Factor Activating NFκB Signaling in Hepatocellular Carcinoma. *PLoS Comput. Biol.* **11,** (2015).

81. Henning, A. N., Haag, J. D., Smits, B. M. G. & Gould, M. N. The Non-coding Mammary Carcinoma Susceptibility Locus, Mcs5c, Regulates Pappa Expression via Age-Specific Chromatin Folding and Allele-Dependent DNA Methylation. *PLoS Genet.* **12,** (2016).

82. Pan, H., Hanada, S., Zhao, J., Mao, L. & Ma, M. Z.-Q. Protein Secretion Is Required for Pregnancy-Associated Plasma Protein-A to Promote Lung Cancer Growth In Vivo. *PLOS ONE* **7,** e48799 (2012).

83. Wang, K.-S., Liu, X.-F. & Aragam, N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr. Res.* **124,** 192–199 (2010).

84. Wang, K.-S. *et al.* Polymorphisms within ASTN2 gene are associated with age at onset of Alzheimer's disease. *J. Neural Transm. Vienna Austria 1996* **122,** 701–708 (2015).

85. Wilson, P. M., Fryer, R. H., Fang, Y. & Hatten, M. E. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. *J. Neurosci. Off. J. Soc. Neurosci.* **30,** 8529–8540 (2010).

86. Chang, H., Cahill, H., Smallwood, P. M., Wang, Y. & Nathans, J. Identification of Astrotactin2 as a Genetic Modifier That Regulates the Global Orientation of Mammalian Hair Follicles. *PLoS Genet.* **11,** e1005532 (2015).

87. Yu, Y. *et al.* Fish TRIM32 functions as a critical antiviral molecule against iridovirus and nodavirus. *Fish Shellfish Immunol.* **60,** 33–43 (2017).

88. Fu, B. *et al.* TRIM32 Senses and Restricts Influenza A Virus by Ubiquitination of PB1 Polymerase. *PLoS Pathog.* **11,** e1004960 (2015).