QTL mapping of intestinal neutrophil variation and inflammation between threespine stickleback populations reveals links to neurodegenerative disease

Emily A. Beck¹, Mark C. Currey¹, Clayton M. Small¹, William A. Cresko¹

¹*Institute of Ecology and Evolution, University of Oregon, Eugene OR, USA*

Corresponding Author:

William A. Cresko

312 Pacific Hall

5289 University of Oregon

Eugene, OR 97403-5289

(541) 346-4779

wcresko@uoregon.edu

Abstract

Host selection is often required to foster beneficial microbial symbionts and suppress deleterious pathogens. In animals, the host immune system is at the center of this relationship. Failed host immune system-microbial interactions can result in a persistent inflammatory response in which the immune system indiscriminately attacks resident microbes, and at times the host cells themselves, leading to diseases such as Ulcerative Colitis, Crohn's Disease, and Psoriasis. Host genetic variation has been linked to both microbiome diversity and to severity of such inflammatory disease states in humans. However, the microbiome and inflammatory states manifest as quantitative traits, which encompass many genes interacting with one another and the environment. The mechanistic relationships among all of these interacting components are still not clear. Developing natural genetic models of host-microbe interactions is therefore fundamental to understanding the complex genetics of these and other diseases. Threespine stickleback (*Gasterosteus aculeatus*) fish are a tractable model for attacking this problem because of abundant population-level genetic and phenotypic variation in the gut inflammatory response. Previous work in our laboratory identified genetically divergent stickleback populations exhibiting differences in intestinal neutrophil activity. We took advantage of this diversity to genetically map variation in an emblematic element of gut inflammation - intestinal neutrophil recruitment - using an F2-intercross mapping framework. We identified three regions of the genome associated with increased intestinal inflammation containing several promising candidate genes. Within these regions we found candidates in the Coagulation/Complement System, NFkB and MAPK pathways along with several genes associated with neurodegenerative diseases commonly accompanying intestinal inflammation as a secondary symptom. These findings highlight the utility of using naturally genetically diverse 'evolutionary mutant models' such as threespine stickleback to better understand interactions among host genetic diversity and microbiome variation in health and disease states.

Introduction

Animals harbor an array of microbes on and in their bodies which perform essential functions that are fundamental to host health (Fraune and Bosch 2010; Relman 2012; McFall-Ngai et al. 2013). Maintaining appropriate host-microbe interactions by facilitating the presence of symbionts and removing pathogens is therefore vital to sustaining health (Bates et al 2006; Blaser and Falkow 2009; Round and Mazmanian 2009; Chung et al. 2012; Jostins et al. 2012). Interactions between the host immune system and resident microbes are at the center of this relationship (Bates et al 2006; Ley et al. 2008; Blaser and Falkow 2009; Round and Mazmanian 2009; Chung et al. 2012; Jostins et al. 2012; Relman 2012; McFall-Ngai et al. 2013). The immune system can promote beneficial microbes that increase host fitness, and failed interactions can result in a persistent inflammatory response, with the immune system chronically responding negatively to resident microbes. This in turn results in diseases such as Ulcerative Colitis and Crohn's Disease (Eckburg and Relman 2007; Emilsson et al. 2008; Graham and Xavier 2013).

The relationship between host immune system and resident microbes is complex. Some microbes cause disease states only in specific host genetic backgrounds or in the presence of other microbes (Casadevall and Pirofski 2000). For example, important work in humans has revealed a strong influence of genetic variation on health outcomes particularly in the context of additional microbiome variation (Dethlefsen et al. 2007; Manolio et al. 2009; Ko et al. 2009; Torkamani et al. 2012; Goodrich et al. 2014). In addition, these host-microbe interactions can be mediated by internal environmental conditions such as stress physiology (Lupp et al. 2007; Alverdy and Luo 2017; Mackenzie et al. 2017) and external conditions such as diet (Hildebrandt et al 2009; Albenberg and Wu 2014; Voreades et al. 2014; Singh et al. 2017). As such, host genetic variation and associated microbiomes can productively be considered quantitative traits.

What is needed are studies that can link quantifiable microbe-induced differences in immune response to host genomic loci and genetic variants. One way to quantify the inflammatory response is through assessment of neutrophils, specialized white blood cells that are recruited during an inflammatory response (Bradley et al. 1982; Renshaw et al. 2006; Kumar and Sharma 2010; Mantovini et al. 2011; Kolaczowska and Kubes 2013). These cells exist throughout the body and are recruited from the blood stream to sites of inflammation, including the gut (Borregard 2010; Fournier and Parkos 2012; Wera et al. 2016). While intestinal neutrophil recruitment often occurs due to the presence of pathogens, it can also occur chronically due to aberrant interactions between the immune system and the gut microbiota (Foell et al. 2002; Wera et al 2016).

Genomic regions that underlie these complex inflammatory phenotypes associated with neutrophil variation can be identified using genetic mapping in model organisms through the use of mutational screens (Musani et al. 2006; Hillhouse et al. 2011; Leach et al. 2012; Uddin et al. 2011; Chen et al. 2016; Barry et al. 2018). Because of the complex interplay of genetics, microbes and environment, it is also essential to develop outbred mutant models tractable for genetic mapping of *natural* genetic variants influencing complex phenotypes such as inflammation (Albertson et al. 2009). Here, we use the threespine stickleback fish (*Gasterosteus aculeatus*) as such an outbred ‘evolutionary mutant model’ (Albertson et al. 2009) to study just such complex disease traits.

This small teleost fish is found throughout the arctic in a wide range of environments including freshwater and oceanic habitats, resulting in exceptional degrees of within -and among- population genetic and phenotypic variation for countless traits (Bell and Foster 1994; Colosimo et al. 2004; Cresko et al. 2004, 2007; Hohenlohe et al. 2010; Glazer et al. 2015). Notably, there are multiple high quality genome assemblies from disparate populations (Jones et al. 2012; Peichel et al. 2017) and the large clutch sizes of stickleback provide ample family sizes for QTL mapping (Colosimo et al. 2004; Cresko et al. 2004; Kimmel et al. 2012; Miller et al. 2014; Glazer et al. 2015; Greenwood et al. 2015; Peichel and Marques 2017). By using threespine stickleback lines originating from genetically diverse populations with distinct ecological and evolutionary histories we are able to map natural genetic variants thus allowing us to identify the types of variants likely underlying this complex phenotype in the human population (Albertson et al. 2009).

Previous work in our laboratory described phenotypic variation between freshwater and oceanic ecotype inflammatory responses, with oceanic individuals responding more robustly to the presence of microbes measured by an increase in intestinal neutrophil accumulation and changes in gene expression (Milligan-Myhre et al. 2016; Small et al. 2017). These findings identified a potential role of host genetic variation on differences in intestinal inflammation and the response to the presence of microbes across populations. We set out to map natural genetic variants associated with differences in intestinal neutrophil density using an F2-intercross genetic mapping study in threespine stickleback. We used these data to identify genomic regions that, when combined with previously published gene expression data from juvenile guts in the parental populations (Small et al. 2017), identified a concordant list of candidate genes involved in host immunity. Surprisingly, we also found several genes with characterized functions in neurodegenerative diseases known to include intestinal inflammation as a secondary symptom. These findings have broader impacts in elucidating roles of natural genetic

variation in chronic intestinal inflammation and provide further evidence of a strong link between intestinal health and systemic inflammation.

Methods

Husbandry and Experimental Design

We generated an F2 mapping cross of threespine stickleback derived from wild caught Alaskan populations previously maintained in the laboratory for at least ten generations. An F1 line was generated by *in vitro* fertilization of parents derived from two distinct Alaskan populations including a male from the freshwater population Boot Lake (N 61.7167, W 149.1167) and a female from the anadromous population Rabbit Slough (N 61.5595, W 149.2583). An F2 family (n = 64) was produced intercrossing F1 siblings. Fertilized eggs were incubated overnight in one-micron filter sterilized antibiotic embryo media containing 100 mg/mL Ampicillin, 50 mg/mL Kanamycin, and 8 mg/mL Amphotericin (Milligan-Myhre et al. 2016; Small et al. 2017). The next day embryos were surface sterilized using 0.003% bleach solution and 0.2% Polyvinylpyrrolidone-iodine (PVP-I) solution (Western Chemical Inc.) following protocols described by Small et al. 2017. Fish were raised in sterile stickleback embryo media until 9 days post fertilization (dpf) when they were moved to a 9.5 liter tank exposing them to the microbes present in the Cresko Lab fish facility. At this time, fish were housed under “summer” conditions of 16 hours of daylight and 8 hours of night where they were fed 2 mL of hatched brine shrimp naupli and fry food (Ziegler AP100 larval food) designed to mimic the diet of wild stickleback. Water temperature was maintained at 20° C with salinity at 4 parts per thousand (PPT) on a recirculating system. At 14 dpf, fish were euthanized with MS222, following IACUC approved methods described by Cresko et al. 2004. Fish were imaged for standard length measurements, tail-clipped, and fixed overnight in 4% paraformaldehyde (PFA) at room temperature then moved to 4°C for long term storage.

MPO Staining and Phenotyping

Whole fish were stained using the Sigma Myeloperoxidase kit (Sigma, 390A-1KT, St Louis, MO, USA), which preferentially stains neutrophils. Fish were then embedded in paraffin, and their bodies were cross-sectioned in 7-micron sections from just posterior to the gill rakers to the urogenital opening. Every 10th section beginning posterior to the gill rakers was imaged. Neutrophils were counted twice in each imaged section. If the counts did not agree, the section was counted a third time. Average neutrophils per section were then calculated and tested for association with standard length and sex using R v3.4.1 (R Core Team 2017).

Statistical Analysis of Phenotypic Variation.

Intestinal neutrophil abundance in 14 days post fertilization (dpf) F2s was distributed roughly normally, with full siblings ranging from less than one neutrophil per section on average to over seven neutrophils per section (Figure 1). This variation was shown to be independent of sex ($t = 1.35$; $P = 0.18$) (Figure S1), as males and females exhibited similar distributions of average intestinal neutrophils. However, neutrophil abundance was shown to be correlated with standard length, with larger fish exhibiting higher average neutrophil density ($R^2 = 0.14$; $P = 0.0013$) (Figure S2). To account for size as a covariate, we calculated residuals from the regression of neutrophil density and standard length, to be included as trait values in our QTL mapping.

DNA Isolation and Sex Determination

Tail samples were flash frozen in liquid nitrogen and stored at -80°C . DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA). DNA was then quantified using the Qubit fluorometer broad range kit. Individual sex phenotypes were determined using PCR amplification of a sex specific region of the genome using the GA1 primer pair (Griffiths 2000), and males were identified by the presence of the male-specific amplicon.

Genotyping of Parents and Progeny

Genomic DNA from each F1 parent and F2 offspring was standardized to 10 ng/uL and digested with the endonuclease SbfI-HF (NEB), and RAD-seq libraries were generated using protocols established by the Cresko Laboratory (Hohenlohe et al. 2010; Baird et al 2008; Etter et al 2011). In some progeny samples, DNA concentrations fell below the 10 ng/uL threshold, but all samples with at least 100 ng of DNA were used. Uniquely barcoded samples were then sequenced in one lane on the Illumina HiSeq 4000 to obtain single end 150 bp reads. To improve coverage, the lane was re-run through a second round of sequencing on the HiSeq 4000. Raw sequence data was demultiplexed by barcode and filtered using the process_radtags program in the Stacks suite v1.48 (Catchen et al. 2011; Catchen et al. 2013). Together these sequencing lanes yielded 799,824,397 reads with 708,390,956 reads retained, averaging 3,873,856 reads retained per individual. Reads were then aligned using GSNAP (Wu and Watanabe 2005) to the stickleback reference genome from Ensembl (version 80), allowing for seven maximum mismatches and a terminal threshold of ten. Genotypes were called using the ref_map pipeline of the Stacks suite.

QTL Mapping

Genotype calls were concatenated in a VCF file generated using the populations package in Stacks (Catchen et al. 2011; Catchen et al. 2013). Filtering was then performed using VCFtools (Danecek et al. 2011). For a marker to be included in the analysis, it was required to have a genotype call in at least 50% of the progeny. MapMaker files were then generated using custom scripts including 18,394 informative markers. QTL mapping was then performed using the *r/QTL scanone* function with Haley-Knott regression (Broman and Sen 2009; Broman et al. 2003). To include growth rate as a covariate, QTL mapping was also performed on residuals calculated from a regression of neutrophils per section on standard length. Genomic regions were further analyzed using the raw neutrophil count data if QTLs were preserved in both the raw and residual data analyses. To account for potential false positives, *r/QTL* was re-run with phenotypes randomly assigned to genotypes.

Functional Assignments of Associated SNPs

To identify differentially expressed genes from the parental populations we used RNA-seq differentially expression analysis data from 14 dpf threespine stickleback guts obtained from Small et al. 2017. To assign function to genes within each of our associated SNP boundaries, we used GeneCard v4.8.1 Build8.

Data Availability

All supplemental files are available on the GSA FigShare portal. Sequencing data are publicly available in the Sequencing Read Archive (SRA) under project ID: SUB879235.

Results

Several genomic regions associated with increased intestinal neutrophil density independent of stickleback size

We identified 18,394 informative SNPs to be used for QTL mapping. Using raw intestinal neutrophil densities (number of neutrophils per section), we detected 13 linkage groups (LG) with QTLs associated with variation in neutrophil density ($LOD > 3$) (Figure 2; Table 1). These included LG1, LG2, LG3, LG4, LG7, LG8, LG9, LG11, LG12, LG13, LG14, LG16, and LG19. To disentangle these findings from standard length and subsequently avoid mapping genes associated with developmental rate or growth, we mapped the residuals of neutrophil number included in a linear model with standard length as a covariate (Figure 2; Table 1). The number of regions of the genome likely associated with intestinal neutrophil recruitment was greatly reduced when mapping the residuals. QTLs retained included pile-ups on LG1, a single SNP on LG4, narrowed regions on LG3 and LG8, and a single pileup on LG12.

In some instances, low coverage or missing genotypes can falsely inflate LOD scores of individual SNPs generating false positives. To evaluate our data for false positives, we randomized the raw phenotype data and the residuals with respect to the F2 genotypes and re-ran r^2/QTL . This method is similar to large scale permutation tests used to designate a significant alpha value, but can also indicate regions where low coverage in specific SNPs are generating false positives. In our original runs, high-LOD regions were concentrated in pileups (Figure 2), indicating the type of associative signal expected from linkage mapping. This is in stark contrast to our randomized dataset, with high-LOD SNPs scattered sparsely across the genome (Figure S3). Most importantly, analysis of the randomized datasets revealed very little overlap with our original runs, providing support that low coverage was not generating false positives (Figure S3). The only overlapping signal between the original and randomized data was the pattern of two pileups on LG1. Requiring further analysis of the SNPs on LG1.

To test if low coverage in a few SNPs could be generating false positives, we assessed phenotype distributions for each genotype in the SNPs with $LOD > 3$. We found false positives on LG1 as expected and additional false positives on LG12. On LG1, four of the five SNPs with $LOD > 3$, neutrophil densities of homozygous individuals did not significantly differ from heterozygous siblings ($t = -0.38$; $P = 0.70$; $df = 48.4$; $t = 1.36$; $P = 0.18$; $df = 49.8$; $t = 1.50$; $P = 0.14$; $df = 42$; $t = -1.74$; $P = 0.09$; $df = 41.5$). In the fifth SNP, however, there initially appeared to be a significant increase in neutrophil density in homozygous siblings ($t = 2.57$; $P = 0.015$; $df = 30.6$). Upon further inspection, however, we discovered one of the F2s that exhibited a high neutrophil density did not have sufficient coverage at this locus to reliably call a genotype. We imputed the genotype of this missing individual based on surrounding loci, and determined this SNP was likely heterozygous. Reanalysis using the imputed genotype resulted in a loss of significance, suggesting the signal on LG1 was in fact a false positive ($t = 1.81$; $P = 0.08$; $df = 31.6$) (Figure S4). Phenotypic distributions consistent with false positives were also identified on LG12 (Figure S5). These markers showed similar patterns to those on LG1 with heterozygous and homozygous siblings exhibiting similar average neutrophil densities ($t = -1.11$; $P = 0.28$; $df = 34.7$, $t = -1.11$; $P = 0.28$; $df = 34.7$ and $t = -1.30$; $P = 0.20$; $df = 43.4$).

At the remaining loci with $LOD > 3$ on LG3, LG4, and LG8, all testing suggested true associations of SNPs with differences in neutrophil density. On LG3 we subsampled five SNPs exhibiting the highest LOD scores in the pile-up. Three SNPs exhibited a similar pattern: with homozygotes exhibiting a significant decrease in neutrophil density compared to their heterozygous siblings (Table 2; Figure 3; Figure S6). The fourth SNP exhibited the opposite pattern with homozygous siblings exhibiting a significant increase in neutrophil density

compared to heterozygous siblings (Table 2; Figure 3; Figure S6). The fifth SNP was generated from two heterozygous parents and therefore was split into three groups: two homozygous groups (AA/TT) and a heterozygous group (AT) (Figure S6). At this locus, fish homozygous for the A allele had a significantly higher average neutrophil density than either other genotypic group while individuals with genotypes AT or TT did not differ significantly (Table 2; Figure 3; Figure S6). On LG4 there were three SNPs encompassing two RAD markers with LOD > 3. At each locus heterozygous siblings averaged a higher neutrophil density than homozygous siblings (Table 2; Figure 3; Figure S6). On LG8 we again subsampled the five SNPs with the highest LOD scores which encompassed four RAD markers. In all cases homozygous siblings had a significantly higher neutrophil density than heterozygous siblings (Table 2; Figure 3; Figure S6).

To demarcate genomic intervals of interest based on our retained QTLs, we defined boundaries using the outside flanking SNPs LOD < 3 (Table S3). In LG3, we identified 11 SNP clusters LOD > 3, including one marker with a score below the initial threshold of LOD > 3. We chose to include this locus, however, as the LOD score of the residual was one of the highest (LOD = 4.13). In LG4 we identified two genomic regions of interest and in LG8 we identified 17 such regions (Table S3).

Immune pathways and disease genes associated with increased intestinal neutrophil density

To ascertain biological relevance of each of our QTLs we used the threespine stickleback genome annotation to compile a complete list of candidate genes associated with each genomic interval. Within these boundaries we assembled a complete gene list and identified those with known functions and if available assigned directionality of differential expression from the parental populations (Small et al. 2017). Within these intervals we identified several gene groups of interest including members of the Coagulation/Complement Cascade, Mitogen-Activated-Protein Kinase (MAPK) pathway, Extracellular Signal-Regulated Kinase (ERK), the Nuclear factor kappa-light-chain-enhancer of activated B cells (NfKB) immune pathway, and the maintenance of tight junctions (Table 3).

The first group of genes were involved in the Coagulation/Complement Cascade, a pathway also enriched for genes differentially expressed between oceanic and freshwater stickleback families (Small et al. 2017). These included *f3*, *c8a*, and *c8b* on LG3 and *par-4* on LG4 (Table 3). On LG3 one marker was located within the 12th exon of *c8a* while on LG4 a marker was located within intron 1 of *par-4* (Table S4). Interestingly, *par-4* is also differentially expressed in

14 dpf guts of our parental populations, exhibiting significant upregulation in oceanic families compared to freshwater (Small et al. 2017).

The second group included members of the ERK signaling and MAPK pathways. Interestingly, the MAPK Pathway was also enriched for genes sensitive to the presence of microbes in freshwater families (Small et al. 2017). Five of these genes were on LG3: *map2k2b*, *epha4b*, *pik3R3*, *angptl3*, and *tcf3a*; six of these genes were on LG8: *borcs8*, *ncan*, *camk4*, *gadd45bb*, *map2k2a* and *tcf3b* (Table 3). None of these genes exhibited differential expression patterns (Small et al. 2017) nor were any markers located within the coding region of these genes, but the identification of two co-orthologous pairs (LG3: *map2k2b* and *tcf3a*; LG8: *map2k2a* and *tcf3b*) in addition to the number of genes in this pathway is highly suggestive of a strong association between MAPK/ ERK signaling and intestinal inflammation.

Another pathway associated with several candidate genes was the NfκB immune pathway. This included *ripk2* - an activator of NFκB - on LG3, *pias4a* and *gadd45bb* on LG8 and *par-4* on LG4 with joint roles in the Coagulation/Complement Cascade and inhibition of the NfκB pathway through direct binding of NFκB1 (Table 3) (Diaz-Meco et al. 1999; Fernandez-Marcos et al 2009; Diaz-Meco and Moscat 2012; Burikhanov et al. 2017). Among these NfκB pathway genes, *par-4* was the only gene differentially expressed in oceanic versus freshwater families, where oceanic families exhibited higher expression (Small et al. 2017).

The last functional group of genes included three genes involved in the formation and maintenance of tight junctions. These included *cldn18* on LG3 and *cldn34* and *zgc:153311* on LG8 (Table 3). Interestingly, *cldn18* and *cldn34* were both upregulated in oceanic compared to freshwater families (Table S4) (Small et al. 2017).

Finally, we identified several gene candidates associated with neurodegenerative diseases that have secondary symptoms relating to intestinal inflammation. This included three genes associated with Amyotrophic Lateral Sclerosis (ALS): *unc13a* on LG3 (Diekstra et al. 2012), *par-4* on LG4 (Xie et al. 2005) and *nek1* on LG8 (Kenna et al. 2016). We also identified three genes associated with Autism Spectrum Disorder (ASD): *tblxr1b* on LG3 and *gabrb3* and *gabrg3* on LG8. Lastly, we identified one disease gene associated with muscular dystrophy, *sgcb*, on LG8.

Discussion

To our knowledge this is the first study in which an F2 intercross mapping framework has been used to identify genomic loci underlying a complex immune trait. The identification of multiple genomic loci significantly associated with variation in neutrophil density suggests

several genes throughout the genome contribute to this complex phenotype. Our results highlight several genomic loci with relatively large effects contributing specifically to intestinal neutrophil recruitment. The persistence of high LOD SNP pile-ups on LG3, LG4, and LG8 after accounting for standard length as a covariate and testing for false positives argue for the biological relevance of the QTLs detected in this analysis. However, the modest size of this single family of F2 progeny ($n = 64$) means that other genomic regions of small effect would likely have gone undetected. Although other genes of small effect with bearing on the observed variation in intestinal neutrophil activity may have been missed, we can be confident in our identification of several genomic intervals on LG3, LG4, and LG8 with strong associations to intestinal neutrophil activity.

The large number of immune genes identified in the associated genomic intervals is not an unexpected result. The Coagulation/Complement Cascade, MAPK, ERK, and NfκB pathways are all essential immune pathways activate early in development and play roles in the regulation of the inflammatory response (Kurosawa et al. 2000; Markiewski and Lambris 2007; Liu et al. 2007; Dev et al. 2011; Simon et al. 2013; Liu et al. 2017). The identification of a variant within *par-4*, a gene with joint roles in regulation of the Coagulation/Complement Cascade and the NfκB pathway is of particular interest as this gene is known to be upregulated in oceanic populations which exhibit a more robust inflammatory response (Small et al. 2017).

The identification of three co-orthologous pairs was also particularly interesting. In the first pair, *c8a* and *c8b* were both found in the same genomic interval with the associated SNP located within *c8a*. These proteins both function as a part of the Complement Cascade, a pathway enriched for differentially expressed genes between oceanic and freshwater stickleback (Small et al. 2017). Together, these proteins initiate membrane penetration and coordinate the formation and insertion of the membrane attack complex (MAC) into the bilayer to facilitate lysis (Bubeck et al. 2011). The other co-orthologous pairs *map2k2a/map2k2b* and *tcf3a/tcf3b* are also excellent candidates for impacting the inflammatory response as *tcf3* has been shown to play a role in the regulation of B cell maturation and *map2k2* is an activator of MAPK. These pairs are of particular interest, however, as they were found on separate linkage groups with *map2k2b/tcf3a* on LG3 and *map2k2a/tcf3b* on LG8.

The identification of several genes involved in the formation and regulation of tight junctions was also intriguing. Tight junctions are extremely important in regulating intestinal permeability and the intestinal immune response, and have been tied to many intestinal diseases including IBD, Celiac Disease, and Type 1 Diabetes (Visser et al. 2010; Castoldi et al. 2015; Lee et al. 2015; Konig et al 2016). Coinciding with this group were two additional disease genes

associated with Celiac Disease and wheat allergies, *rgs1* and *scp2a*, which contained three markers within intronic regions (Table 3; Table S4).

Lastly, the identification of variants within neurodegenerative disease genes provides potential links between intestinal and neurological health. Our findings included variants associated with known disease genes impacting ALS, Autism, and Muscular Dystrophy. All of these diseases are of particular interest to those studying intestinal inflammation, as individuals who exhibit them often report symptoms consistent with colitis and other types of intestinal inflammation at greater rates; and in many cases targeted treatments of the microbiome have been successful in alleviating or slowing progression of symptoms (Nowak et al. 1982; Bellini et al. 2006; Kaneko and Hachiya 2006; Fang 2016; Lo Cascio et al. 2016; Rowin et al. 2017; Hughes et al. 2018; Opazo et al. 2018; Patusco and Ziegler 2018; Spielman et al. 2018; Wright et al. 2018). How the genetics of these complex diseases are tied to intestinal health is still an unresolved problem and requires further mapping of inflammatory phenotypes, but these targets provide a strong starting point to investigate broader implications on intestinal inflammation on systemic health.

Conclusions

This study provides a strong example of the power of threespine stickleback as a model for mapping natural variants contributing to genetically complex phenotypes relevant to human disease. To expand upon these findings, we can use the stickleback system to map other related immune system phenotypes and expand studies of inflammation to other tissue types. These findings additionally provide potential targets for functional testing using CRISPR-Cas9 genome editing to connect systems-level genetic links between neuro-muscular disorders and intestinal health. Because of the amenability of stickleback for gnotobiotic studies, these genetic approaches will be particularly useful to manipulate both host genes and microbiomes simultaneously to perform functional tests not possible in other organisms. Our findings have broader impacts in elucidating the complex roles of natural genetic variation in chronic intestinal inflammation and provide further evidence of a strong link between intestinal health and systemic health.

References

- Albenberg, L.G. and G.D. Wu. (2014). Diet and the Intestinal Microbiome: Associations, Functions, and Implications for Health and Disease. *Gastroenterol* 146:1564-1572.
- Albertson, R. C., Cresko, W. A., Detrich H. W. 3rd, Postlethwait, J. H. (2009) Evolutionary mutant models for human disease. *Trends Genet* 25: 74 -81.
- Alverdy, J. C. and J.N. Luo. (2017). The Influence of Host Stress on the Mechanism of Infection: Lost Microbiomes, Emergent Pathobiomes, and the Role of Interkingdom Signaling. *Front Microbiol* 8:322.
- Baird, N. A., Etter, P. D. Atwood, T. C., Currey, M. C., Shiver, A. L. et al. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3: e3376.
- Barry, J. D. Fagny, M., Paulson, J. N., Aerts, H. J. W. L. Platig, J. et al. (2018) Histopathological Image QTL Discovery of Immune Infiltration Variants. *iScience* 5: 80-89.
- Bates, J. M., Mittge, E., Kuhlman, J., Baden, K. N., Cheesman, S. E. et al. (2006) Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Dev Biol* 297: 374-386.
- Bell, M.A. and S.A. Foster. (1994). *The Evolutionary Biology of the Threespine Stickleback*. Oxford University Press.
- Bellini, M., Biagi, S., Stasi C., Costa, F., Mumolo, M. G. et al. (2006) Gastrointestinal manifestations in myotonic muscular dystrophy. *World J Gastroenterol* 12: 1821-1828.
- Blaser, M. J. and S. Falkow. (2009) What are the consequences of the disappearing human microbiota? *Nat Rev Microbiol* 7: 887-94.
- Borregaard N. (2010) Neutrophils, from marrow to microbes. *Immunity* 33: 657-670.
- Bradley, P. P., Priebat, D. A., Christensen, R. D., Rothstein, G. (1982) Measurement of cutaneous inflammation: estimation of neutrophil content with an enzyme marker. *J Invest Dermatol* 78: 206-209.
- Broman, K. W., Wu, H. Sen, S., Churchill, G. A. (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889-890.
- Broman, K. W. and S. Sen. (2009) *A guide to QTL mapping with R/qtl*. Springer. <https://rqtl.org/book>
- Bubeck, D., Roversi, P., Donev, R., Morgan, B.P., Llorca, O. et al. (2011). Structure of human complement C8, a precursor to membrane attack. *J Mol Biol* 405: 325-330.
- Buikhanov, R. Hebbar, N., Noothi, S. K., Shukla, N., Sledziona, J. et al. (2017) Chloroquine-Inducible Par-4 Secretion is Essential for Tumor Apoptosis and Inhibition of Metastasis. *Cell Rep* 18: 508-519.

- Casadevall, A. and L. Pirofski (2000) Host-Pathogen Interactions: Basic Concepts of Microbial Commensalism, Colonization, Infection, and Disease. *Infect Immun* 68: 6511-6518.
- Castoldi, A., Favero de Aguiar, C., Moraes-Viera, P. M., Olsen Saraiva Camara, N. (2015) They Must Hold Tight: Junction Proteins, Microbiota And Immunity In Intestinal Mucosa. *Curr Protein Pept Sci* 16: 655-671.
- Catchen, J. Amores, A., Hohenlohe, P. A., Cresko, W. A., Postlethwait, J. H. (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3 (Bethesda)* 1: 171-182.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amore, A., Cresko W. A. (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22: 3124-3140.
- Chen, L., Ge. B., Casale, F. P., Vasquez, L., Kwan, T. et al. (2016) Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* 167: 1398-1414.
- Chung, H., Pamp, S. J., Hill, J.A., Surana, N. K., Edelman, S. M. et al. (2012) Gut Immune Maturation Depends on Colonization with a Host-Specific Microbiota. *Cell* 149: 1578-593.
- Colosimo, P. F., Peichel C. L. Nereng, K., Blackman B. K., Shapiro, M. D. et al. (2004) The Genetic Architecture of Parallel Armor Plate Reduction in Threespine sticklebacks. *PLoS Biol* 2: E109.
- Cresko, W. A. Amores, A., Wilson, C., Murphy, J., Currey, M. et al. (2004) Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc Natl Acad Sci U S A* 101: 6050- 6055.
- Danecek, P., Auton A., Abecasis G., Albers C. A., Banks, E. et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156-2158.
- Dethlefsen, L., McFall-Ngai, M., Relman, D. A. (2007) An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* 449: 811-818.
- Dev, A., Iyer, S., Razani, B., Cheng, G. (2011). NF- κ B and innate immunity. *Curr Top Microbiol Immunol* 349:11-43.
- Diaz-Meco, M. T., Lallena, M. J., Monjas, A., Frutos, S., Moscat. J. (1999) Inactivation of the inhibitory κ B protein kinase/nuclear factor κ B pathway by Par-4 expression potentiates tumor necrosis factor α -induced apoptosis. *J Biol Chem* 274: 19606-19612.
- Diaz-Meco, M. T. and J. Moscat. (2012) The atypical PKCs in inflammation: NF- κ B and beyond. *Immunol Rev* 246: 154-167.
- Diekstra, F. P., van Vught, P. W., van Rheenen, W., Koppers, M., Pasterkamp, R. J. et al. (2012) UNC13A is a modifier of survival in amyotrophic lateral sclerosis. *Neurobiol Aging* 33: 630.e3-8.
- Eckburg, P. B. and D. A. Relman (2007) The role of microbes in Crohn's disease. *Clin Infect Dis* 44: 256-62.

- Emilsson V., Thorleifsson, G., Zhang, B., Leonardson A. S., Zink, F. et al. (2008) Genetics of gene expression and its effect on disease. *Nature* 452: 423-428.
- Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., Cresko, W. A. (2011) SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. *Methods Mol Biol* 772: 157-178.
- Fang, X. (2016) Potential role of gut microbiota and tissue barriers in Parkinson's disease and amyotrophic lateral sclerosis. *Int J Neurosci* 126: 771-776.
- Fernandez-Marcos, P. J., Abu-Baker, S., Joshi, J., Galvez, A., Castilla, E. A. (2009) Simultaneous inactivation of Par-4 and PTEN in vivo leads to synergistic NF-kB activation and invasive prostate carcinoma. *Proc Natl Acad Sci U S A* 106: 12962-12967.
- Foell D., Kucharzik, T., Kraft, M., Vogl, T., Sorg, C. et al. (2003) Neutrophil derived human S100A12 (EN-RAGE) is strongly expressed during chronic active inflammatory bowel disease. *Gut* 52: 847-853.
- Fournier, B. M. and C. A. Parkos. (2012) The role of neutrophils during intestinal inflammation. *Mucosal Immunol* 5: 354-366.
- Fraune, S. and T. C. Bosch. (2010) Why bacteria matter in animal development and evolution. *Bioassays* 32: 571-80.
- Gasch, A. P. Payseur, B. A., Pool. J. E. (2016). The power of natural variation for model organism biology. *Trends Genet* 32: 147-154.
- Glazer, A. M., Killingbeck, E. E., Mitros, T. Rokhsar, D. S., Miller, C. T. (2015) Genome Assembly Improvement and Mapping Convergent Evolutionary Traits in Sticklebacks with Genotyping-by-Sequencing. *G3 (Bethesda)* 5: 1463-1472.
- Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L. Koren, O. et al. (2014) Human genetics shape the gut microbiome. *Cell* 159: 789-799.
- Graham, D. B. and R. J. Xavier (2013). From genetics of inflammatory bowel disease towards mechanistic insights. *Trends Immunol* 34: 371-8.
- Greenwood, A. K., Ardekani, R., McCann, S. R., Dubin, M. E., Sullivan, A. et al. (2015) Genetic mapping of natural variation in schooling tendency in threespine stickleback. *G3 (Bethesda)* 5: 761-769.
- Griffiths R. (2000) DNA sex identification in the three-spined stickleback. *J Fish Biol* 57: 1331-1334.
- Hildebrandt, M.A., Hoffman, C., Sherrill-Mix, S.A., Keilbaugh, S. A., Hamady, M. et al. (2009). High-Fat Diet Determines the Composition of the Murine Gut Microbiome Independently of Obesity. *Gastroenterol* 137: 1716-1724.
- Hillhouse, A. E., Myles, M. H., Taylor, J. F., Bryda, E. C., Franklin C. L. (2011) Quantitative trait loci in a bacterially induced model of inflammatory bowel disease. *Mam Genome* 22: 544-555.

- Hohenloe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. et al. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 6: e1000862.
- Hughes, H. K., Rose, D., Ashwood, P. (2018) The Gut Microbiota and Dysbiosis in Autism Spectrum Disorders. *Curr Neurol Neurosci Rep* 18: 81.
- Jones, F.C., Grabherr, M.G, Chan, Y.F., Russel, P, Mauceli, E., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484: 55-61.
- Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H. McGovern, D. P. et al. (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491: 119-124.
- Kaneko, K. and N. S. Hachiya. (2006) Hypothesis: gut as source of motor neuron toxin in the development of ALS. *Med Hypotheses* 66:438-9
- Kenna, K. P., van Doormaal, P. T., Dekker, A. M., Ticozzi, N., Kenna B. J. et al. (2016) NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat Genet* 48: 1037-1042.
- Kimmel, C. B., Ullmann, B., Currey, M., Hohenlohe, P. A., Cresko, W. A. (2012) Developmental dissociation in morphological evolution of the stickleback opercle. *Evol Dev* 14: 326-337.
- Ko, D. C., Shukla, K. P., Fong, C., Wasnick, M., Brittnacher, M. J., et al. (2009) A genome-wide in vitro bacterial-infection screen reveals human variation in the host response associated with inflammatory disease. *Am J Hum Genet* 85: 214–227.
- Kolaczowska E. and P. Kubes (2013) Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol* 13: 159-175.
- Konig, J., Wells, J., Cani, P. D., Garcia-Rodenas, C. L. MacDonald, T. et al. (2016) Human Intestinal Barrier Function in Health and Disease. *Clin Transl Gastronertol* 7: e196.
- Kumar, V. and A. Sharma (2010) Neutrophils: Cinderella of innate immune system. *Int Immunopharmacol* 10: 1325-1334.
- Kurosawa, M., Numazawa, S., Tani, Y., Yoshida, T. (2000). ERK signaling mediates the induction of inflammatory cytokines by bufalin in human monocytic cells. *Am J Cell Phys* 278: 500-508.
- Leach, R. J., O'Neill, R. G., Fitzpatrick, J. L., Williams, J. L. Glass, E. J. (2012) Quantitative Trait Loci Associated with the Immune Response to a Bovine Respiratory Syncytial Virus Vaccine. *PLoS ONE* 7: e33526.
- Lee, S.H. (2015) Intestinal Permeability Regulation by Tight Junction: Implication on Inflammatory Bowel Diseases. *Intest Res* 13: 11-18.
- Lescak, E. A. and K. C. Milligan-Mhyre. (2017) Teleosts as Model Organisms To Understand Host Microbe interactions. *J Bacteriol* 199: e00868-16.

Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., Gordon, J. I. (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* 6: 776-788.

Liu, T., Zhang, L., Joo, D., Sun, S. (2017). NF- κ B signaling in inflammation. *Sig Trans Target Ther* 2: 17023.

Liu, Y., Shepherd, E.G., Nelin, L.D. (2007). MAPK phosphatases – regulating the immune response. *Nat Rev Immunol* 7:202-212.

Lo Cascio, C. M., Goetze, O., Latshang, T. D., Bluemel, S Frauenfelder, T. et al. (2016) Gastrointestinal Dysfunction in Patients with Duchenne Muscular Dystrophy. *PLoS ONE* 11: e0163779.

Lupp, C. Robertson, M.L., Wickham, M.E., Sekirov, I., Champion, O.L. et al. (2007). Article host-mediated inflammation disrupts the intestinal microbiota and promotes overgrowth of *Enterobacteriaceae*. *Cell Host Microbe* 2: 119-129.

Mackenzie, B. W., Waite, D. W., Hoggard, M., Gouglas, R.G., Tayloe, M.W., et al. (2017). Bacterial community collapse: a meta-analysis of the sinonasal microbiota in chronic rhinosinusitis. *Environ Microbiol* 19: 381-392.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A. et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747-753.

Mantovani, A., Cassatella, M. A., Costantini, C., Jaillon, S. (2011) Neutrophils in the activation and regulation of innate and adaptive immunity. *Nat Rev Immunol* 11: 519-531.

Markiewski, M. M. and J.D. Lambris. (2007). The Role of Complement in Inflammatory Diseases From Behind the Scenes into the Spotlight. *Am J Pathol* 171: 715-727.

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C., Carey, H.V. Domazet-Loso, T. et al. (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* 110: 3229-3236.

Miller, C. T., Glazer, A. M., Summers, B. R. Clackman, B. K., Norman, A. R. et al. (2014) Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait Loci. *Genetics* 197: 405-420.

Milligan-Mhyre, K., Small, C. M., Mittge, E. K., Agarwal, M. Currey, M. (2016) Innate immune responses to gut microbiota differ between oceanic and freshwater threespine stickleback populations. *Dis Model Mech* 9: 187-198.

Musani, S. K., Zhang, H., Hsu, H., Yi, N. Gorman, B. S. et al. (2006) Principal component analysis of quantitative trait loci for immune response to adenovirus in mice. *Hereditas* 143: 189-197.

Nowak, T. V., Ionasescu, V., Anuras A. (1982) Gastrointestinal manifestations of the muscular dystrophies. *Gastroenterol* 82: 800-810.

- Opazo, M. C., Ortega-Rocha, E. M., Coronado-Arrazola, I., Bonifaz, L. C., Boudin, H. et al. (2018) Intestinal Microbiota Influences Non-intestinal Related Autoimmune Diseases. *Front Microbiol* 9: 432.
- Patusco, R. and J. Ziegler. (2018) Role of Probiotics in Managing Gastrointestinal Dysfunction in Children with Autism Spectrum Disorder: An Update for Practitioners. *Adv Nutr* 9: 637-650.
- Peichel, C. L. and D. A. Marques. (2017) The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos Trans R Soc Lond B Biol Sci* 372: 20150486.
- Peichel, C.L., Sullivan, S. T., Lichko, I, and M.A. White. (2017) Improvement of the Threespine Stickleback Genome Using Hi-C-Based Proximity- Guided Assembly. *J of Hered.* 108:693-700.
- Relman, D. A. (2012) The human microbiome: ecosystem resilience and health. *Nutr Rev* 70: S2-9.
- Renshaw, S. A., Loynes, C. A., Trushell, D. M., Elworthy, S., Ingham, P. W. et al. (2006) A transgenic zebrafish model of neutrophilic inflammation. *Blood* 108: 3976-3978.
- Round, J. L. and S. K. Mazmanian (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 9: 313-23.
- Rowin, J.m Xia, Y., Jung, B., Sun, J. (2017) Gut inflammation and dysbiosis in human motor neuron disease. *Physiol Rep* 5: e13443.
- Simon, J., Arthur, C., Ley, S.C. (2013). Mitogen-activated protein kinases in innate immunity. *Nat Rev Immunol* 13: 679-692.
- Singh, R. K., Chang, H., Yan, D. Lee, K. M., Ucmak, D. et al. (2017). Influence of diet on the gut microbiome and implications for human health. *L Trans Med* 15: 73.
- Small, C. M., Milligan-Mhyre, K., Bassham, S., Guillemin, K. Cresko, W. A. (2017) Host Genotype and Microbiota Contribute Asymmetrically to Transcriptional Variation in the Threespine Stickleback Gut. *Genome Biol Evol* 9: 504-520.
- Spielman, L. J., Gibson, D. L., Klergis, A. (2018) Unhealthy gut, unhealthy brain: The role of the intestinal microbiota in neurodegenerative diseases. *Neurochem Int* 120: 149-163.
- Torkamani, A., Pham, P., Libiger, O., Bansal, V., Zhang, G. et al. (2012) Clinical Implications of human population differences in genome-wide rates of functional genotypes. *Front Genet* 3: 211.
- Uddin, M. J., Cinar, M. U., Groe-Brinkhaus, C., Tesfaye, D., Tholen, E. et al. (2011) Mapping quantitative trait loci for innate immune response in the pig. *Int J Immunogenet* 38: 121-131.
- Visser, J., Rozing, J., Sapone, A., Lammers, K. Fasano, A. (2010) Tight Junctions, Intestinal Permeability, and Autoimmunity Celiac Disease and Type 1 Diabetes Paradigms. *Ann N Y Acad Sci* 1165: 195-205.
- Voreades, N., Kozil, A., Weir, T. L. (2014). Diet and the development of the human intestinal microbiome. *Front Microbiol* 5:494.

Wera, O. Lancellotti, P., Oury, C. (2016) The Dual Role of Neutrophils in Inflammatory Bowel Diseases. J Clin Med 5: 118.

Wright, M. L., Fournier C., Houser, M. C., Tansey, M., Glass, J et al. (2018) Potential Role of the Gut Microbiome in ALS: A Systematic Review. Biol Res Nurs 20: 513-521.

Wu, T. D. and C. K. Watanabe. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21: 1859-1875.

Xie, J., Awad, K. S., Guo, Q. (2005) RNAi knockdown of Par-4 inhibits neurosynaptic degeneration in ALS-linked mice. J Neurochem 92: 59-71.

Competing Interests

The authors declare no competing interests.

Acknowledgements

We would like to thank Susan Bassham and Angel Amores for helpful discussions and comments on the manuscript. We would also like to thank Micaela Burns, Emily Niebergall, Kayla Sharp, and Nathalie Verhoeven for help with imaging and morphometric measurements.

Author Contributions

The study was designed by E.A.B., C.M.S., and W.A.C. Data Analyses were conducted by E.A.B and M.C.C. The initial manuscript was written by E.A.B with all authors contributing to subsequent versions.

Funding

This work was supported by the National Institute of General Medical Sciences and the National Center for Research Resources of the National Institutes of Health under award numbers P50GM098911 to (W.A.C, K. Guillemin et al.), R24RR032670 to (W.A.C) and a National Institute of Health NRSA fellowship F32GM122419 to E.A.B.

Figure Legends

Figure 1. Phenotypic Variation of Intestinal Neutrophil Recruitment. **(A)** Gut sections of a 14 dpf fish with a low average neutrophil count per section. **(B)** Gut sections of a 14 dpf fish with a high average neutrophil count per section. **(C)** F2 Phenotypic distribution of variation of intestinal neutrophil recruitment.

Figure 2. QTL Maps of Average Neutrophil Counts. **(A)** Raw neutrophil count data **(B)** Residual data including growth rate as a covariate. Linkage Groups alternate in color Black/Grey. Horizontal line indicates a LOD cutoff of 3.

Figure 3. Phenotype Distributions by Genotype. **(A)** Zoomed in view of Manhattan plots of raw QTL data with individual points displaying the highest LOD scores colored based on patterns observed in boxplots. **(B)** Corresponding boxplots to high LOD score SNPs. Relationships between homozygotes and heterozygotes from multiple SNPs are concatenated into single boxplots when they exhibit the same pattern.

Supplemental Figure 1. Phenotypic Distribution by Sex.

Supplemental Figure 2. Phenotypic Distribution by Standard Length.

Supplemental Figure 3. QTL Maps of Randomized Neutrophil Counts **(A)** Randomized raw neutrophil count data **(B)** Randomized residual data including growth rate as a covariate. Linkage Groups alternate in color Black/Grey. Horizontal line indicates a LOD cutoff of 3.

Supplemental Figure 4. Phenotype Distributions of all analyzed SNPs by Genotype in LG1. Open shapes correspond to boxplots below. Red data points indicate genotype calls missing from the SNP generating a false positive.

Supplemental Figure 5. Phenotype Distributions of all analyzed SNPs by Genotype on LG12.

Supplemental Figure 6. Phenotype Distributions of all analyzed SNPs on (A) LG3 (B) LG4 (C) LG8.

[illegible]

Table 2. Phenotypic Distribution summary by SNP.

Location	Mean Neutrophil Density Homozygote #1	Mean Neutrophil Density Heterozygote	Mean Neutrophil Density Homozygote #2	t^a	df^b	P
LG3						
SNP1	3.04	4.52	NA	-3.15	29.01	0.004
SNP2	3.13	4.18	NA	-2.67	49.70	0.010
SNP3	3.06	4.66	NA	-2.99	23.00	0.006
SNP4	4.18	3.13	NA	2.72	50.06	0.009
SNP5	3.162723 (TT)	3.057886 (AT)	5.147527 (AA)	TT/AT: 0.27049 TT/AA: -3.5077 AT/AA: -4.2347	TT/AT: 20.16 TT/AA: 20.68 AT/AA: 15.08	TT/AT: 0.79 TT/AA: 0.002 AT/AA: 0.0007
LG4						
SNP1	2.83	4.13	NA	-3.90	51.83	0.0003
SNP2	3.12	4.24	NA	-2.77	40.67	0.008
SNP3	3.12	4.24	NA	-2.77	40.67	0.008
LG8						
SNP1	4.16	2.84	NA	3.87	56.93	0.0003
SNP2	4.16	2.84	NA	3.87	56.93	0.0003
SNP3	4.30	3.06	NA	3.28	46.92	0.002
SNP4	4.20	2.78	NA	4.02	52.98	0.0002
SNP5	4.07	2.96	NA	3.09	53.90	0.003

a- Welch Two Sample t-test statistic

b- Degrees of freedom

Table 3. Candidate Gene Summary.

Group	Ensembl ID	Gene	Distance from SNP* (bp)	Immune Pathways	Disease Associations
LG3	ENSGACG00000014756	<i>f3</i>	37,324	Coagulation Cascade	-
LG3	ENSGACG00000015273	<i>map2k2b</i>	84,711	MAPK/ERK	-
LG3	ENSGACG00000015282	<i>tcf3a</i>	58,000	MAPK/ERK	-
LG3	ENSGACG00000015301	<i>unc13a</i>	15,909	-	ALS
LG3	ENSGACG00000015386	<i>epha4b</i>	25,437	MAPK/ERK	-
LG3	ENSGACG00000015719	<i>cldn18</i>	10,244	Tight Junctions	-
LG3	ENSGACG00000016028	<i>ripk2</i>	6,127	NFkB	-
LG3	ENSGACG00000016066	<i>tblxr1b</i>	16,774	-	Autism
LG3	ENSGACG00000016121	<i>rgs1</i>	1,892	-	Celiac Disease
LG3	ENSGACG00000016159	<i>pik3R3</i>	63,634	MAPK/ERK	-
LG3	ENSGACG00000016189	<i>scp2a</i>	0	-	Wheat Allergy
LG3	ENSGACG00000016212	<i>angptl3</i>	5,349	MAPK/ERK	-
LG3	ENSGACG00000016323	<i>c8b</i>	972	Complement Cascade	-
LG3	ENSGACG00000016338	<i>c8a</i>	0	Complement Cascade	-
LG4	ENSGACG00000020015	<i>par-4</i>	0	Coagulation Cascade; NFkB	ALS
LG8	ENSGACG00000012681	<i>pias4a</i>	68,402	NFkB	-
LG8	ENSGACG00000012686	<i>map2k2a</i>	58,202	MAPK/ERK	-
LG8	ENSGACG00000012737	<i>tcf3b</i>	5,424	MAPK/ERK	-
LG8	ENSGACG00000013097	<i>borcs8</i>	19,084	MAPK/ERK	-
LG8	ENSGACG00000013123	<i>ncan</i>	0	MAPK/ERK	-
LG8	ENSGACG00000013599	<i>camk4</i>	6,656	MAPK/ERK	-
LG8	ENSGACG00000013618	<i>gadd45bb</i>	5,900	MAPK/ERK; NFkB	-
LG8	ENSGACG00000013753	<i>gabrb3</i>	21,271	-	Autism
LG8	ENSGACG00000013771	<i>gabrg3</i>	0	-	Autism
LG8	ENSGACG00000014026	<i>cldn34</i>	17,633	Tight Junctions	-
LG8	ENSGACG00000014029	<i>zgc:153311</i>	15,831	Tight Junctions	-
LG8	ENSGACG00000014071	<i>nek1</i>	80,789	-	ALS
LG8	ENSGACG00000014199	<i>sgcb</i>	0	-	Muscular Dystrophy

* indicates closest SNP LOD > 3