Identifying genetic loci and spleen gene coexpression networks underlying immunophenotypes in BXD recombinant inbred mice

Rachel M. Lynch,1 Sudhir Naswa,2 Gary L. Rogers, Jr.,2 Stephen A. Kania,3 Suchita Das,1 Elissa J. Chesler,1,4 Arnold M. Saxton,5 Michael A. Langston,2 and Brynn H. Voy1,5

1Systems Genetics Group, Oak Ridge National Laboratory, Oak Ridge; Departments of 2Electrical Engineering and Computer Science and 3Comparative Medicine, University of Tennessee, Knoxville, Tennessee; 4The Jackson Laboratory, Bar Harbor, Maine; and 5Department of Animal Science, University of Tennessee, Knoxville, Tennessee

Submitted 12 January 2010; accepted in final form 22 February 2010

Lynch RM, Naswa S, Rogers GL Jr, Kania SA, Das S, Chesler EJ, Saxton AM, Langston MA, Voy BH. Identifying genetic loci and spleen gene coexpression networks underlying immunophenotypes in BXD recombinant inbred mice. Physiol Genomics 41: 244–253, 2010. First published February 23, 2010; doi:10.1152/physiolgenomics.00020.2010.—The immune system plays a pivotal role in the susceptibility to and progression of a variety of diseases. Due to a strong genetic basis, heritable differences in immune function may contribute to differential disease susceptibility between individuals. Genetic reference populations, such as the BXD (C57BL/6J × DBA/2J) panel of recombinant inbred (RI) mouse strains, provides unique models through which to integrate baseline phenotypes in healthy individuals with heritable risk for disease because of the ability to combine data collected from these populations across both multiple studies and time. We performed basic immunophenotyping (e.g., percentage of circulating B and T lymphocytes and CD4+ and CD8+ T cell subpopulations) in peripheral blood of healthy mice from 41 BXD RI strains to define the immunophenotypic variation in this strain panel and to characterize the genetic architecture that underlies these traits. Significant QTL models that explained the majority (50–77%) of phenotypic variance were derived for each trait and for the T:B cell and CD4+:CD8− ratios. Combining QTL mapping with spleen gene expression data uncovered two quantitative trait transcripts, Ptprk and Acp1, as candidates for heritable differences in the relative abundance of helper and cytotoxic T cells. These data will be valuable in extracting genetic correlates of the immune system in the BXD panel. In addition, they will be a useful resource for prospective, phenotype-driven model selection to test hypotheses about differential disease or environmental susceptibility between individuals with baseline differences in the composition of the immune system.

quantitative trait loci; quantitative trait transcript; eQTL; CD4:CD8 ratio; T:B cell ratio; T lymphocytes; B lymphocytes; Ptprk; Acp1

SYSTEMS GENETICS TAKES a top-down approach to disease susceptibility by seeking to identify relationships between genetic variants, intermediate molecular, biochemical and cellular pathways, and overlying systems-level phenotypes from large-scale molecular and phenotypic analyses. Typically, putative interconnections are built through correlational analyses of diverse data types collected across genetically heterogeneous populations. These data are often obtained using genetic reference populations, i.e., populations that are genetically stable and thus reproducible, allowing data integration across time and from diverse studies and creating the possibility to uncover novel relationships among genes, pathways, and diseases. A number of genetic reference populations exist for mouse, the largest of which is the BXD (C57BL/6J × DBA/2J) recombinant inbred (RI) strain panel, consisting of 81 extant strains for which genotype data are publicly available (56). As the depth of phenotyping for a reference panel like the BXD strain set increases, so does the ability to interconnect physiological systems through genetic correlation of phenotypes. Such discoveries can be valuable for determining the molecular basis for a phenotype, for identifying biomarkers for disease processes, and for elucidating interconnections between apparently divergent physiological systems.

The burgeoning evidence that inflammation either initiates or fuels a wide variety of diseases and pathologies suggests that immune system components are likely to emerge in many systems-level networks of disease susceptibility. Disorders not traditionally linked with the immune system such as obesity and insulin resistance are now causally linked to inflammatory processes and mobilization of immune cells (36, 54, 78). While the abundance of specific lymphocyte subpopulations is altered by numerous environmental factors such as infection and diet (5, 73, 86), these traits are also under tight genetic control (2, 25, 41). The involvement of immune processes in myriad diseases, many of which have a genetic risk, coupled with a strong genetic basis for immune function, raises the possibility that genetic variation in immune phenotypes per se may contribute to differential disease susceptibility between individuals. Supportive of this concept is a recent report by Dendrou et al. (28), using the Cambridge BioResource, which is the human equivalent of a genetic reference population consisting of ~5,000 healthy, genotyped individuals living near Cambridge, UK, who agreed to be studied repeatedly over time. Dendrou and colleagues linked a specific T cell phenotype - relative expression of CD25 on the surface of CD4+ memory T cells - with a haplotype previously shown to confer protection from type I diabetes. Understanding the genetic basis of specific immunophenotypes (IPs) may therefore be valuable in understanding susceptibilities and/or progression of diseases, such as multiple sclerosis and rheumatoid arthritis, which are characterized by dysregulation or imbalances of distinct immune cell populations (26, 35, 63, 68).

As a first step toward determining the heritable differences in IPs that may predict outcomes of disease and environmental exposures, we profiled the abundance of major lymphocyte subpopulations (CD79a, CD3e, CD4+, and CD8+) in peripheral blood of healthy, unperturbed mice from the BXD strain panel. These data were integrated with existing genotype data in quantitative trait loci (QTL) mapping to characterize the
genetic architecture that underlies these traits. They were further combined with microarray data we collected from spleens of the same strain panel to highlight potential candidate genes within significant QTL. Graph algorithms were used to identify gene expression networks that correlate with and are potential mediators of IPs, and are thus potential targets for environmental stimuli that act on the immune system. Herein we present evidence that peripheral IPs are under significant genetic control in the BXD population and describe genes and coexpression networks linking genetic variation to immunophenotypic diversity.

**METHODS**

**Animals.** C57BL/6J, DBA/2J, and BXD RI stocks from strains 6–42 were obtained from The Jackson Laboratory (Bar Harbor, ME). BXD RI stocks from strains 43–100 were obtained from Dr. Lu Lu and Dr. Robert Williams from the University of Tennessee Health Science Center (UTHSC, Memphis, TN). BXD RI lines were housed and propagated in the specific pathogen-free (SPP) Russell Vivarium at Oak Ridge National Laboratory (ORNL). Mice received irradiated Purina Diet #5083 and chlorinated water ad libitum. The housing conditions were maintained at 70 ± 2°F and 40–60% humidity. A total of 45 BXD strains were used for spleen expression profiling, immunophenotyping, or both. This subset of the BXD panel was chosen from the set of strains maintained at ORNL based on consistent breeding performance and to represent a balanced combination of the original BXD strains developed at The Jackson Laboratory (71, 72) and the advanced intercross strains developed at the UTHSC (56). Between 10 and 12 wk of age, mice were killed by cervical dislocation and either blood was collected for immunophenotyping or the spleens were harvested for RNA expression profiling. All studies were approved by the Animal Care and Use Committee at Oak Ridge National Laboratory.

**Immunophenotyping.** Flow cytometry was used for the immunophenotyping of male and female mice (average of four mice/sex/strain) from 41 BXD strains and the parental strains. Blood was collected by retro-orbital sinus puncture into EDTA tubes, and red blood cells were lysed using lysis buffer (Sigma-Aldrich, St. Louis, MO). Following centrifugation, the white blood cell pellet was suspended in buffer (PBS, 0.2% sodium azide, 0.02% heat-inactivated FBS) and divided into four aliquots for each set of monoclonal antibodies and a blank negative control. Lymphocytes were stained with the appropriate antibody or antibodies for 45 min at 4°C. The negative control was incubated with PBS. One tube was dual-stained with anti-CD3 (PE, clone 17A2) and anti-CD79b (FITC, clone HM79b). Another tube was dual-stained with anti-CD4 (PE, clone H129.19) and anti-CD8a (FITC, clone 53-6.7). The remaining tube was stained with anti-MHC class II (R-PE, clone NIMR-4). Antibodies were purchased from BD Biosciences (Franklin Lakes, NJ), except MHC class II-RPE, which was purchased from Southern Biotech (Birmingham, AL). Following incubation, the samples were centrifuged and suspended in PBS. All samples were stored on ice in the dark until analyzed by flow cytometry. At least 10,000 cells per sample were analyzed using a Beckman Coulter Epics XL flow cytometer (Brea, CA). Data were analyzed using EXPo32 ADC Software (Beckman Coulter). Lymphocytes were gated for analysis based on forward and side scattering profiles. The IPs measured included the proportion of circulating T cells (%CD3), B cells (%CD79), CD4+ T cells (%CD4), CD8+ T cells (%CD8), as well as the median expression of major histocompatibility complex II (MHCII median) on MHCII+ lymphocytes (%MHCII).

**Identification of immunophenotype QTL.** Flow cytometric data were analyzed for quality based on efficient staining of lymphocytes and within-individual consistency (e.g., %CD79 approximately equaling %MHCII and sum of %CD4 plus %CD8 approximately equaling %CD3). Only high-quality immunophenotype data were used for further analysis, resulting in an average of 3.3 males and 3.2 females per strain for each IP. T cell to B cell ratio, CD4+ to CD8+ ratio, and MHC II median fluorescence were normalized using natural log transformation (i.e., LN T:B, LN CD4:CD8, LN MHCII).

**QTL analysis.** QTL analysis was performed using genotype data obtained from GeneNetwork (http://www.genenetwork.org/dbdoc/BXDDeno.html). This database contains nearly 3,800 informative single-nucleotide polymorphisms (SNPs) and microsatellite markers originally reported by Shifman et al. (62) that have been re-aligned with National Center for Biotechnology Information (NCBI) Build 36. QTL for each IP were identified using the QTL package (16) in R (http://www.r-project.org). The multiple imputation method of Sen and Churchill (60) was used to perform single-QTL genome-wide scans. Genomewide significance thresholds were calculated based on 1,000 permutations (24). The cut-off P values for significant and suggestive loci were P = 0.05 and P = 0.63, respectively (42). The ±1 logarithm of the odds ratio (LOD) support intervals for each QTL were calculated using the lodint function in R/QTL. Multiple-QTL modeling was performed using stepwise linear regression in SAS (SAS Institute, Cary, NC); a P value of 0.05 was used as the threshold for terms to remain in the final model.

**Expression profiling.** Transcriptome profiling was performed in spleens from an independent set of mice representing 38 BXD strains (34 of which were immunophenotyped). Total spleen RNA was isolated from spleens stabilized in RNAlater (Sigma-Aldrich) using RNeasy Mini Kits (Qiagen, Valencia, CA). RNA concentration and quality were assessed using Experion RNA StdSens Chips on the Experion system (Bio-Rad, Hercules, CA). Each BXD sample profiled consisted of a pool of equal amounts of RNA from either two males or two females per strain. Expression profiling was performed by Genome Quebec (Montreal, Canada) using the Mouse WG-6 v.1.1 beadChip on the Illumina platform (San Diego, CA). Six strains were analyzed per chip. The data were normalized using variance stabilizing transformation followed by robust spline normalization using the R/lumi package (30) in Bioconductor (32). Raw and normalized microarray data have been uploaded to NCBI’s Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/projects/geo; accession GSE19935) according to MIAME standards (14). Expression data, along with all IP data, are also available through GeneNetwork (http://www.genenetwork.org).

**Transcriptome map and expression QTL analysis.** Of the 34,492 probes in the Illumina arrays, 11,445 transcripts demonstrated variable expression across the panel (coefficient of variation >0.01) and were used for expression QTL (eQTL) analysis. QTL Reaper (http://www.genenetwork.org/qtlreaper.html) was used to identify the maximum likelihood ratio statistic and a permitted P value (1,000 permutations) for each transcript. QTL Reaper performs Haley-Knott regression for QTL analysis, with an adaptive permutation by transcript that runs an increased number of permutations for those traits with significant results to ensure precise P value estimation at the low end of the P distribution. This method is fast and sufficient for high-density marker maps as are available for the BXD RI lines. At a P value threshold of 0.05 over the entire array, 1,881 transcripts were identified using the QTL package (16) in R (http://www.r-project.org). The multiple imputation method of Sen and Churchill (60) was used to perform single-QTL genome-wide scans. The QTL Reaper analysis was performed using genotype data obtained from GeneNetwork (http://www.genenetwork.org/dbdoc/BXDDeno.html). This database contains nearly 3,800 informative single-nucleotide polymorphisms (SNPs) and microsatellite markers originally reported by Shifman et al. (62) that have been re-aligned with National Center for Biotechnology Information (NCBI) Build 36. QTL for each IP were identified using the QTL package (16) in R (http://www.r-project.org). The multiple imputation method of Sen and Churchill (60) was used to perform single-QTL genome-wide scans. Genomewide significance thresholds were calculated based on 1,000 permutations (24). The cut-off P values for significant and suggestive loci were P = 0.05 and P = 0.63, respectively (42). The ±1 logarithm of the odds ratio (LOD) support intervals for each QTL were calculated using the lodint function in R/QTL. Multiple-QTL modeling was performed using stepwise linear regression in SAS (SAS Institute, Cary, NC); a P value of 0.05 was used as the threshold for terms to remain in the final model.

**Statistical modeling and graph algorithms.** MGI (http://www.informatics.jax.org) was used to extract genes located within the ±1 LOD support intervals for the two significant QTLs identified for the CD4:CD8 immunophenotype. Stepwise linear regression in SAS was...
used to model the CD4:CD8 ratio using the expression of these genes. Graph algorithms were performed as described previously (76), and all source codes are available from the authors. Graphs were created from microarray data by computing all pair-wise Pearson correlations between expressed transcripts and then filtering the matrix to retain only statistically significant correlations based on a false discovery rate of 5% (q-value < 0.05, Ref. 67). Maximal cliques were extracted and anchored cliques identified as previously described (76). A bipartite graph was created with one partition being the expressed transcripts and the second partition being a set of five IPs (CD3%, LN T:B, CD4%, CD8%, LN CD4:CD8). Pearson correlations were computed between each possible pairing of transcript expression and immunophenotype. Edges in the bipartite graph were filtered to retain correlations of interest (P < 0.001), and maximal bicliques were then extracted. Gene ontology (GO) enrichment analysis was performed using DAVID (37); Benjamini-Hochberg false discovery rate-corrected P values are reported (8).

RESULTS

We began by profiling a panel of IPs in male and female C57BL/6J and DBA/2J mice to define baseline differences in the two BXD parental strains. Peripheral blood was analyzed for proportion of circulating T cells (CD3+), B cells (CD79+), T helper cells (CD4+), cytotoxic T cells (CD8+), and expression of major histocompatibility complex II (MHCII) on MHCII+ lymphocytes using flow cytometry. As shown in Table 1, the parental strains differ significantly in each of these traits (P < 0.05) except for CD3+ cells (P = 0.184) and CD8+ T cells (P = 0.064). C57BL/6J mice demonstrate a higher percentage of circulating B cells and CD8+ T cells and lower levels of CD4+ T cells compared with DBA/2J mice, which is consistent with previous reports (21, 52, 74).

The same panel of IPs was profiled across a set of 41 BXD strains to establish the immunophenotypic diversity in this RI panel and to model the genetic regulation of each trait. ANOVA indicated a significant effect of strain on each IP (P < 0.0001), manifested as broad ranges for each of the traits across the strains. For example, the percentage of T lymphocytes varies over fivefold (10.3–56.1%), while the percentages of CD4+ and CD8+ lymphocytes vary over sixfold (5.6–35.6 and 4.4–26.4%, respectively) (Fig. 1, Supplemental Table S1). These ranges relative to those of the parentals illustrate the genetic complexity of the traits and are within the range of the phenotypic diversity found in a survey of 32 standard inbred strains of diverse origin (10, 57). There is also a significant strain*sex interaction effect for each immunophenotype (P < 0.05), but the effect is much smaller (F statistic < 30%) than that of the main strain effects. Thus, overall strain effects were modeled for each immunophenotype.

QTL analysis was performed on each IP as well as the T:B and CD4:CD8 ratios to identify overlapping and unique loci associated with each trait. We began with MHC II density as a reference trait with a defined genetic basis on Chromosome (Chr) 17, a region in which the BXD parental strains carry different haplotypes (C57BL/6J carries the H-2^b haplotype and DBA/2J carries the H-2^k). Not surprisingly, the genotype at the H2 locus is the largest factor in explaining median expression of MHCII on MHCII-expressing lymphocytes (Supplemental Fig. S1A). The locus itself accounts for 57% of the variance within the BXD panel (LOD = 7.52, P < 0.001), with the DBA/2J genotype increasing MHCII expression. A suggestive secondary locus, Chr 10 @ 114Mb, accounts for an additional 4.5% of the variance. The genetic basis for variation in the remaining IPs was modeled by first performing single model genome-wide scans (Supplemental Fig. S1) and then using suggestive and significant loci from those scans in multilocus regression to allow for additive and interactive contributions of multiple loci. A 9 Mb region on Chr 17 spanning the H2 locus was identified as either suggestive or significant for single model scans of each trait except for the CD4:CD8 ratio (Fig. 2). The majority of variance (50.1–77.5%) for each trait is explained by models incorporating the Chr 17 region in combination with at least three additional loci (Table 2). T and B cells were measured as a percentage of total gated lymphocytes and collectively represent the majority of this population, making these two measurements highly inversely correlated. Accordingly, the QTL models for each trait are very similar. The relative abundance between these two cell types (natural log transformed T:B ratio, or LN T:B) was mapped to identify loci that may be involved in lymphocyte maturation. The T:B phenotype is partially explained by interactions between loci on Chrs 3 and 7 and between the Chr 3 locus and H2 locus. In combination with the additive effects of the H2 locus and an additional region on Chr 11, this model explains 77.5% of the variation in the T:B ratio. The only locus important in modeling the T:B ratio that was not identified for either lymphocyte population separately (i.e., %CD3 and %CD79) is the locus on Chr 7. The complete set of significant and suggestive QTLs for each immunophenotype is depicted in Fig. 2.

Genetic control of T cell abundance was further probed by mapping loci that contribute to variance in the major T cell subpopulations, CD4+ and CD8+ T cells. While QTL analysis identified the H2 locus as contributing to both %CD4 and %CD8 (LOD of 4.07 and 3.79, respectively), the remaining QTLs were unique to the particular T cell subpopulations (Fig. 2). A multilocus model consisting of three loci (on Chrs 10, 12, and the H2 locus) explains 50.1% of the variance in %CD4, while a model consisting of a locus on Chr 3 with the H2 locus explains 56.6% of the variance in %CD8. The ratio of CD4+ to CD8+ cells was used in QTL analysis after normalization using natural log transformation (LN CD4:CD8). Two loci (Chr 10 @ 31Mb, LOD = 4.24 and Chr 12 @ 31Mb, LOD = 4.28) were identified as significant, independent contributors to this trait (Supplemental Fig. S1C, Table 2). At both loci, the DBA/2J allele shifts the ratio in favor of CD4+ cells, consistent

Table 1. Differences in peripheral blood immunophenotypes in C57BL/6J and DBA/2J mice

<table>
<thead>
<tr>
<th>Immunophenotype</th>
<th>C57BL/6J</th>
<th>DBA/2J</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>% CD3 lymphocytes</td>
<td>32.06 ± 1.03</td>
<td>35.60 ± 2.21</td>
<td>0.1840</td>
</tr>
<tr>
<td>% CD79 lymphocytes</td>
<td>61.40 ± 0.98</td>
<td>53.86 ± 2.73</td>
<td>0.0352*</td>
</tr>
<tr>
<td>LN T:B</td>
<td>-0.65 ± 0.03</td>
<td>-0.42 ± 0.09</td>
<td>0.0392*</td>
</tr>
<tr>
<td>% CD4 lymphocytes</td>
<td>23.84 ± 1.64</td>
<td>31.56 ± 2.41</td>
<td>0.0190</td>
</tr>
<tr>
<td>% CD8 lymphocytes</td>
<td>7.98 ± 0.56</td>
<td>6.50 ± 0.47</td>
<td>0.0636</td>
</tr>
<tr>
<td>LN CD4:CD8</td>
<td>1.10 ± 0.10</td>
<td>1.58 ± 0.07</td>
<td>0.0002†</td>
</tr>
<tr>
<td>% MHCII lymphocytes</td>
<td>61.62 ± 1.87</td>
<td>51.22 ± 2.64</td>
<td>0.0194*</td>
</tr>
<tr>
<td>LN MHCII density</td>
<td>5.33 ± 0.16</td>
<td>4.31 ± 0.08</td>
<td>0.0003†</td>
</tr>
</tbody>
</table>

Data shown as means ± SE of 5–9 mice per strain. P value indicates significance of strain effect (*P < 0.05, †P < 0.01).
with the increased abundance of CD4$^+$ cells in the DBA/2J parents (Table 1). These two loci explain nearly 54% of the variation in the CD4:CD8 ratio. Interestingly, neither region is implicated in control of either T cell subtype as analyzed independently, suggesting that these two regions contain genetic variation that contributes to a differentiation process in which one cell type is retained at the expense of the other.

eQTL profiling has emerged as a means to identify loci linked directly or indirectly to regulation of gene expression (15, 29, 38, 59). We performed eQTL mapping using spleen microarray data to determine whether the QTL regions identified for IP traits also harbored trans-eQTL bands, loci that are linked to expression of multiple genes and could thus be implicated in mechanistic control of the trait(s) through coordinated transcriptional regulation. Spleen was chosen because it contains abundant levels of both B and T lymphocytes and contributes to multiple aspects of immune function. Two trans-bands exceeded the maximal size of 18 transcripts obtained by permutation testing, consisting of 42 and 30 transcripts located on Chrs 4 (at 139.0 Mb) and 12 (at 15.8 Mb), respectively (Supplemental Fig. S2). Neither trans-band colocalized with IP QTLs, nor were the transcripts associated with each trans-band enriched for functions suggestive of IP regulation. Both bands did, however, contain an abundance of genes involved in cell cycle, cell division and DNA replication, which may have general relevance for heritable regulation of gene expression in the BXD panel.

We further exploited the microarray data to identify potential IP candidate genes by determining if expression of one or more genes within the QTL intervals were correlated with the underlying trait(s), a strategy that has been used successfully for other traits (7, 45, 55). We identified quantitative trait transcripts (QTTs), transcripts whose expression is correlated with a phenotype (55), for the T:B and CD4:CD8 ratios. We focused on these two traits because they capture relative abundance of multiple cell types and have significant QTLs. Pearson’s correlation coefficients were calculated between expression levels of genes residing within the ± 1 LOD support intervals for each QTL and the immunophenotype data. Of the 517 genes located within these intervals for T:B ratio, 88 showed signifi-
correlated with T:B but reside on other chromosomes, the strongest of which are mitofusin 1 \((r = 0.626, P < 0.0001)\) and phospholipase D1 \((r = -0.563, P = 0.0005)\), both of which are within the QTL on Chr 3. Mitofusins are mitochondrial fusion proteins that have recently been linked to the innate antiviral defense system \((82)\). Phospholipase D1 was shown to be critical for coordination of inflammatory signaling through TNF-\(\alpha\) in leukocytes \((61)\).

Similar analysis of the QTL regions for the CD4:CD8 phenotype highlights a smaller set of potential candidate genes for this trait in the BXD panel. Of the 115 genes located within the QTL intervals for CD4:CD8 ratio, only 9 (7.8\%) were significantly correlated with the CD4:CD8 phenotype. Of these, the most highly correlated QTT \((r = 0.575; P = 0.0004)\) is the transcript for protein tyrosine phosphatase, receptor type, K \((Ptprk)\), a phosphatase expressed in spleen and other tissues \((81)\). Loss of Ptprk due to a spontaneous deletion in Long Evans Cinnamon (LEC) rats was recently shown to underlie the deficiency in CD4\(^+\) T cells in this model \((40)\). Conversely, our data link increased expression of Ptprk with elevated levels of CD4\(^+\) cells. Stepwise linear regression of Ptprk and the eight other QTL interval genes correlated with the CD4:CD8 ratio was used to estimate the amount of trait variance explained by expression of these genes. A model containing Ptprk along with acid phosphatase 1 \((Acp1)\); also known as low molecular weight protein tyrosine phosphatase and laminin B-1 \((Lamb1–1)\) explains 61\% of variance in the CD4:CD8 phenotype \((P < 0.0001)\). Like Ptprk, Acp1 is also a strong candidate gene for heritable variation in CD4:CD8 in the BXD panel. Polymorphisms in the human ACP1 gene have been correlated with susceptibility to a number of inflammatory and autoimmune disorders such as type I diabetes, allergy, and atherosclerosis, all disorders in which CD4 and/or CD8 cells are implicated in pathogenesis \((9, 18, 49)\). Lamb1–1 has not been linked specifically to immune function but is widely expressed in spleen \((46)\).

---

**Table 2. Summary of final QTL models for BXD immunophenotypes**

<table>
<thead>
<tr>
<th>Immunophenotype</th>
<th>Final Model</th>
<th>Variance Explained, %</th>
<th>QTLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>%CD3</td>
<td>Loc1*2 + Loc2</td>
<td>58.9</td>
<td>Marker rs13477026 Chr 3 Pos 27 LOD 3.72</td>
</tr>
<tr>
<td>%CD79</td>
<td>Loc1 + Loc2 + Loc3</td>
<td>67.2</td>
<td>rs13482963 Chr 17 Pos 35 LOD 5.28</td>
</tr>
<tr>
<td>LN T:B</td>
<td>Loc1<em>2 + Loc1</em>4 + Loc3 + Loc4</td>
<td>77.5</td>
<td>rs134877026 Chr 3 Pos 27 LOD 4.17</td>
</tr>
<tr>
<td>%CD4</td>
<td>Loc1*2 + Loc3</td>
<td>50.1</td>
<td>rs13481119 Chr 11 Pos 79 LOD 2.54</td>
</tr>
<tr>
<td>%CD8</td>
<td>Loc1*2 + Loc2</td>
<td>56.6</td>
<td>rs13482947 Chr 17 Pos 31 LOD 5.41</td>
</tr>
<tr>
<td>LN CD4:CD8</td>
<td>Loc1 + Loc2</td>
<td>53.8</td>
<td>rs13479274 Chr 7 Pos 58 LOD 2.64</td>
</tr>
<tr>
<td>LN MHC median</td>
<td>Loc1 + Loc2</td>
<td>61.5</td>
<td>rs13482947 Chr 17 Pos 31 LOD 7.52</td>
</tr>
</tbody>
</table>

Models were analyzed using significant and suggestive quantitative trait loci (QTLs) from single locus genome-wide scans. Representative single nucleotide polymorphism markers for each locus in the final models are reported; genome locations are according to NCBI Build 36. LOD, logarithm of odds scores based on single locus models. *Significant QTLs; all others are suggestive.

---

**Fig. 2. Summary of BXD immunophenotype (IP) quantitative trait loci (QTLs).** The ranges indicate the \pm 1\ logarithm of the odds ratio (LOD) support intervals for each significant and suggestive QTL, in Mb. IPs are %CD3 lymphocytes (CD3), %CD79 lymphocytes (CD79), LN of CD3:CD79 (T:B), %CD4 lymphocytes (CD4), %CD8 lymphocytes (CD8), LN of CD4:CD8 (CD4:CD8), and LN of MHCII median expression on MHCII \(^+\) lymphocytes (MHC). *Significant QTL \((P < 0.05)\) in single-locus genome-wide scan. IP labels in black are QTLs that are significant in the final model \((P < 0.05)\); gray IP labels are not significant in final model.
Our group has developed a number of graph algorithms based on the extraction of cliques and other dense subgraphs to identify putative gene coexpression networks from large-scale omics data (6, 11, 22, 43, 76, 84). Here we use the concept of anchored clique to extract networks of genes coexpressed with Acp1 and Ptprk that may provide insight into the mechanisms linking each gene to the CD4:CD8 phenotype. Both genes encode phosphatases and have been linked through genetic association studies to a number of inflammatory conditions, but relatively little is known about the cellular pathways in which each gene functions. The largest clique containing relatively little is known about the cellular pathways in which each gene functions. The largest clique containing Acp1 in a graph thresholded at q-value = 0.05 (with corresponding \( P = 0.0026 \)) consisted of a total of 500 transcripts. The correlations of genes within this clique range from \( 1d = 0.550 \) to 0.917. GO enrichment revealed that this Acp1 coexpression network is highly enriched for genes involved in cell cycle (\( P = 2.0e-22 \), Benjamini = 1.1e-18), cell division (\( P = 1.6e-20 \), Benjamini = 4.2e-17), and DNA replication (\( P = 2.4e-18 \), Benjamini = 3.2e-15). To insure that this high degree of GO enrichment was not somehow an artifact of our specific dataset, we identified genetic correlates of Acp1 expression in an independent spleen array dataset from 24 BXD strains within GeneNetwork (27). As with our data, Acp1 expression was highly significantly correlated with genes involved in cell cycle and related processes (e.g., GO term M phase, Benjamini = 1.13e-18). The cell cycle enrichment in our Acp1 anchored clique is also represented in the KEGG pathway for cell cycle control (\( P = 3.3e-12 \), Benjamini = 6.5e-10). Interestingly, of the 20 genes included in the KEGG pathway and coexpressed with Acp1, 18 are regulated by phosphorylation. These results suggest that abundance of the Acp1 transcript per se may regulate phosphorylation status of cell cycle targets, either directly or indirectly, in the spleen. Similar analysis was performed for the maximal anchored clique of 135 genes containing Ptprk, but significant GO enrichment was not observed.

One of our long-term interests is to identify gene coexpression networks linked to immune function in healthy individuals and to determine how these networks are perturbed by environmental factors that promote inflammation and/or alter immune function. To visualize the intersection between gene networks associated with IPs, we used a biclique algorithm to identify the largest set of transcripts in which each member is significantly correlated with each other and one or more phenotypes. Pearson correlations were computed between each transcript and five immunophenotype measurements (%CD4, %CD3, LN T:B, %CD8, and LN CD4:CD8), and the matrix was thresholded to retain only those correlations most likely to symbolize true relationships (\( P < 0.001 \)), leaving a total of 218 transcripts that are significantly correlated with one or more IPs. The relationship between gene expression and IPs are represented in Fig. 3, and as expected significant overlap exists between the sets of genes associated with each trait. The members of each biclique are listed in Supplemental Table S2. For example, Notch4 is correlated with %CD3, LN T:B, %CD4, and %CD8. Notch4 is part of the highly conserved Notch family which play important roles in lymphocyte lineage commitment (70) and other cell-fate decisions (3). Specifically, there is evidence to support the role of Notch4 in the differentiation and expansion of hematopoietic stem cells and in lymphomagenesis (75, 83). These data provide a starting point from which to test the impact of specific environmental variables on networks of genes linked to IPs of interest.

DISCUSSION

Systems genetics enables the detection of QTLs and the identification of putative candidate genes for further testing and validation. In parallel, it produces phenomic data that can be mined in the context of the system by integrating it with all of the other multiscale and diverse data types obtained from the same population. At a total of 79 strains, the BXD RI strain panel is the largest inbred mouse RI panel, and one for which an abundance of genomic resources now exist (56). We characterized the range of variation and the genetic architecture of IPs in the BXD RI panel to produce a baseline profile of the immune system that can be integrated into systems genetics studies with this population. These data are relevant for genetic susceptibility to the plethora of environmental factors and disorders that invoke the immune system, e.g., ionizing radiation exposure and diet-induced obesity.

We found that genetic models explain the majority of variation in each of the IPs, which is consistent with reports from humans and other species that lymphocyte subpopulations in healthy individuals are under strong genetic control. Three of the IPs (proportions of B, CD4+ and CD8+ lymphocytes in peripheral blood) were previously profiled in 22 of the same BXD lines included in our study (21). Two of the three IPs are significantly correlated with our data (%B lymphocytes, \( r^2 = 0.530, P = 0.010 \); %CD8 lymphocytes, \( r^2 = 0.437, P = 0.041 \)) despite the passage of several years between studies and the use of independent BXD breeding colonies and institutions, highlighting the genetic stability of these phenotypes. The QTLs identified for these three traits differ between the two studies, likely because the Chen and Harrison (21) study relied on 35 BXD strains from the original BXD panel (strains 1–42)
and a limited number of genetic markers, while ours was balanced between the original BXD set and the advanced recombinant inbred set produced by Peirce et al. (56). Recent polymorphisms in this population (62) have been shown to influence the location and direction of QTL effects in these populations (58). In addition, a significant increase in the number of informative genetic markers are now available for the BXD panel and were used herein, which also may contribute to the discrepancy in QTL positions. Our reported QTL intervals (and those for many other traits mapped using the BXD panel) will likely be further refined with upcoming availability of genotype data across a panel of 580,000 SNPs, data that will soon be available through the GeneNetwork website (R. W. Williams, personal communication).

Identifying QTLs is relatively straightforward and rapid when using genetic reference panels such as BXD for which genotype data are readily available; cloning the causative polymorphism and confirming its role in phenotypic determination is a much more elusive target and one that we have not attempted. However, by combining QTL mapping with gene expression data, we have winnowed the list of potential candidate genes for traits of interest to a manageable number for further study. This approach produced two particularly compelling candidate genes - Ptprk and Acp1 - within the two significant QTLs associated with CD4:CD8, a trait used clinically as a marker for the prognosis of human immunodeficiency virus (HIV) (47), rheumatoid arthritis (31), and other diseases. Ptprk encodes a receptor-type protein-tyrosine phosphatase with no specific properties that link it to lymphocyte development. However, a potential role for Ptprk in CD4 T cell development was serendipitously uncovered in studies of the LEC rat, a model of Wilson’s disease (due to a mutation in the copper transporting ATPase gene) that also had been noted to be deficient in CD4+ T cells (1). Recently, two groups (4, 40) used linkage analysis to identify a deletion containing Ptprk in LEC rats and confirmed that loss of Ptprk is responsible for defective CD4+ T cell development. Clinically, deletions of the chromosomal region containing PTTPRK (chromosome 6q22–23) are frequently present in high-grade non-Hodgkin’s lymphoma (53, 85), and loss of this region is also predictive of poor prognosis in CNS lymphoma (20), implicating this phosphatase in oncogenesis of the immune system. We used coexpression analysis in an attempt to gain insight into Ptprk function in spleen, based on the concept of genetic correlation. However, many of the transcripts with which Ptprk is most tightly coexpressed are un- or poorly annotated, and no specific functions showed statistical enrichment. On the other hand, Acp1, the second candidate gene of interest for the CD4:CD8 phenotype, is part of a large clique of 500 transcripts that is highly enriched in functions related to cell cycle. Acp1, also known as low molecular weight protein tyrosine phosphatase, regulates phosphorylation status of a number of proliferative signaling molecules (64) and is upregulated in several types of cancers (48, 50). Genetic screens in humans link polymorphisms in or near the ACP1 locus to a variety of inflammatory diseases including allergy, asthma, and obesity (12), and Acp1 is involved in activation, adhesion, and differentiation of T cells (13, 33). The QTLs containing Acp1 and Ptprk reside on separate chromosomes that showed significant independent and additive linkage with the CD4:CD8 phenotype, suggesting that genes within each locus may interact to affect CD4:CD8 ratio. At the molecular level Ptprk positively regulates the protein tyrosine kinase Src (77), which in turn phosphorylates and activates Acp1 (17, 69). Whether these events occur in the same cell type, and in one relevant to T cell development, is unknown but is worthy of further experimentation. It is worth noting that both Acp1 and Ptprk have been linked through genetic screens to colon cancer susceptibility (65, 66). Germane to our overarching interest in radiation sensitivity is the fact that both Acp1 and Ptprk have been shown to be altered by radiation exposure (34, 79).

Although Ptprk and Acp1 emerge as attractive candidates for the CD4:CD8 phenotype, a potential limitation of our study is that the tissue (spleen) used to highlight these QTLs is not the site of T lymphocyte lineage commitment and selection, processes thought to primarily occur within the thymus (19). However, it has been suggested that there are compensatory adjustments that occur within peripheral immune organs (i.e., spleen and lymph nodes) and alters the T cell peripheral population (52). Therefore, it is possible that the expression levels of Ptprk and Acp1 in the thymus and spleen are regulated through similar mechanisms, or that the genetic polymorphisms that cause variation between strains exert similar effects in both tissues. If so, the relationship between trait values and expression levels would be predicted, even though the specific tissue profiled is not the primary tissue involved in the process of interest. Alternatively, the causative polymorphisms within the QTL intervals, whether in Ptprk and Acp1 or other genes or regulatory elements, may act on processes that occur in the periphery, such as T cell proliferation. However, the mice used in this study were not exposed to any known immune challenges and were maintained in an SPF facility, and we would expect peripheral proliferation of T cells to be minimal in these animals. It is also possible that, despite convergent evidence to support their contribution to CD4:CD8, the causative polymorphism for variation in this trait do not act through either Acp1 or Ptprk.

Genetic correlations between IPs in the BXD panel and disease susceptibility would be enriched by more detailed characterization of T cell subpopulations beyond the classification as either Th (CD4+) or Tc (CD8+). For example, additional surface markers could be used to further classify CD4+ cells as Th1, Th2, Th17, regulatory T (Treg), follicular helper T (Tfh), γδ T cells, etc. Each of these cell types play important roles in host defense and autoimmune diseases, and understanding the genetic basis of T cell subpopulation distributions would be invaluable in elucidating susceptibilities to T cell-mediated disorders, such as rheumatoid arthritis (39, 44, 80).

One of the advantages of using BXD strains as a reference population is the wealth of genotypic, phenotypic, and gene expression information available for the panel, much of which is freely available through the online database GeneNetwork. For example, we transiently uploaded our data into GeneNetwork to scan for genetic correlation between our IPs and other traits measured on the BXD population. This ad hoc analysis revealed an association between the baseline T:B ratio profiled in our study and outcomes of Chlamydia psittaci exposure as measured by Miyairi and colleagues (51). Among BXD strains that persist 30 days postinfection, both spleen weight and pathogen load in spleen (GeneNetwork records 11025 and 11026) are significantly correlated (r = −0.811, P < 0.001, and r = −0.728, P = 0.003, respectively) with the T:B
phenotype. The opportunity for such integrative analyses provided by the use of genetic reference populations such as the BXD panel highlights the strengths of systems genetics, namely the ability to assay genetic susceptibility to disease or the environment in the context of the healthy state through the stable genetic basis of the population. As the BXD strain continues to be phenotyped, the ability to connect seemingly disparate phenotypes grows proportionally. The Collaborative Cross, an idealized genetic reference population, should also be widely available within the next few years for expanded systems genetics studies (23).

In summary, we have characterized the genetic architecture of a set of basic but informative IPs in the BXD panel. We have uncovered potential candidate genes that contribute to genetic variation in the relative abundance of helper and cytotoxic T cells, and follow-on studies to test the roles of both Ptprk and variation in the relative abundance of helper and cytotoxic T uncovered potential candidate genes that contribute to genetic systems genetics studies (23).

Acknowledgments

We thank K. T. Cain for breeding the BXD mice at ORNL, Ginger D. Shaw for collecting blood and spleen samples, Vivek M. Philip for assistance with QTL mapping, and Dianne J. Trent for immunophenotyping.

Grants

The project was supported by the Low Dose Radiation Research Program of the Office of Biological and Environmental Research, Office of Science, Department of Energy ERKP650 and ERKP804.

Disclosures

No conflicts of interest are declared by the authors.

References


