

SYSTEMS GENETICS CAN PROVIDE NEW INSIGHTS TO IMMUNE REGULATION AND AUTOIMMUNITY

By

Grant Morahan^a Violet Peeva^a, Munish Mehta^a, & Rob Williams^b

a. Centre for Diabetes Research,
The Western Australian Institute for Medical Research
University of Western Australia.
Medical Research Foundation Building,
50 Murray St, Perth, WA 6000 Australia

b. Department of Anatomy and Neurobiology
Developmental Genetics, Department of Pediatrics
University of Tennessee Health Science Center
855 Monroe Avenue
Memphis TN 38163 USA

Corresponding author: G. Morahan,
Email: gem@waimr.uwa.edu.au,
Phone: (+61-8) 9224-0206
FAX: (+61-8) 9224-0322

ABSTRACT

“Systems Genetics” detects variation in phenotypic traits and integrates this with underlying genetic variation. A powerful application of system genetics is analyzing effects of genome-wide genetic variants on transcriptome-wide variation in gene expression. We see systems genetics as a new powerful technology which will empower research in immunology and in other disciplines. Here, we present preliminary analyses of systems genetics approaches to defining genetic interactions within the thymus, which is the key site for T lymphocyte development and imposition of immunological tolerance.

Keywords: systems genetics, thymus, microarray, gene networks, complex genetic diseases

Introduction

Most scientists still see genetics as Mendel envisioned it, whereby an individual gene encodes a specific product/protein/phenotype. However, many of the major public health problems society faces – such as diabetes, cardiovascular diseases, cancers, etc – while they are genetically based, do not follow this paradigm. This is certainly true of autoimmune diseases, which arise from complex interactions between multiple genes and environmental factor(s). Understanding such non-Mendelian, complex genetic diseases is the challenge for what may be considered as second generation genetics. Beyond this, the field we have dubbed “Systems Genetics” offers the opportunity to define interacting clusters and networks of genes within a tissue or cell population. This work has been pioneered by Rob Williams and colleagues (1-3) and taken up other groups (e.g. 4-6). Previously termed “genetical genomics” we consider “Systems Genetics” a better description (7).

Microarray technology is central to Systems Genetics, but its use is two orders of magnitude beyond conventional microarray studies. The first quantum difference is that, instead of examining tissues from a single source, tissues from as many as 100 related strains are examined. The second level is that the transcriptomic data is integrated with the underlying genotypic data of each of the strains. Systems Genetics makes it possible to define networks and clusters of co-regulated transcripts; to correlate phenotype with genotype; and to map genes which affect the expression levels of other genes.

Two resources are required for this work to be possible. First, one needs a collection of related, genetically-defined individuals that can be used as a source of tissues. The "systems" aspect of systems genetics suggests testing and correlating a range of phenotypes from a group of related individuals; this is obviously much more difficult if one does not have an immortal mapping population. Recombinant inbred (RI) mouse strains are ideal for this purpose. Each strain is derived from two progenitor strains and has a mosaic genome comprised of chromosome segments inherited from one of the two parental strains. Perhaps the best studied RI set is the BXD series, derived from C57BL/6 and DBA/2 strains (7). An original panel of 26 strains produced by Benjamin Taylor (7) is now augmented by an additional 63 strains (8). Each strain has inherited a different pattern of genes, so differences between strains can be mapped on the basis of allele sharing by similar strains. Crucially, all

mice within a strain are essentially genetically identical, so results may be accumulated from different tissues at different time points and under differing stimulation conditions; results may be confirmed subsequently and by other researchers. Both quantitative and qualitative traits may be mapped using RI strains (7). While it is possible to use F2, backcross or even heterogeneous stock animals for this purpose, these are obviously limited as each has a unique genotype and can offer definition of only a limited number of phenotypes.

Secondly, a new set of bioinformatic tools is required. A very powerful resource for systems genetics analyses has been developed under the *GeneNetwork* umbrella, which incorporates the *WebQTL* programs (1, 9). WebQTL, as the name suggests, is available online (www.genenetwork.org), allowing researchers worldwide to access a range of accumulated data.

Here, we will describe some applications of Systems Genetics to develop new insights into genetic regulation of immunologically significant genes, using as an example the first immune-related organ for which data is currently available: our unpublished data on thymus.

Methods

Mice. C57BL/6 mice and BXD strains (7, 8) were obtained from colonies maintained at the University of Tennessee. Thymus tissues were dissected from at least 5 mice from each of 26 different strains.

Microarrays. RNA was purified from duplicate pools of 2-3 thymus samples per strain. cRNA samples were prepared and hybridized to Illumina Ref 8 v1.1 arrays according to the manufacturer's instructions. Data was cleaned by background subtraction and rank invariant normalization using the Illumina Beadstudio package.

Data analyses. The normalized data were exported in 'CSV' format and uploaded into the WebQTL Database Schema by an interpreter tool developed in Python. The data and the WebQTL programs (1, 9) were installed on an Apple X-RAID server (Apple Computers, Cupertino, CA). Access to WebQTL is available via the Gene Network web site, www.genenetwork.org.

Results

In this paper, we will illustrate some of the features of systems genetics as currently enabled by WebQTL. We will provide as an example the expression in thymus of genes selected to

complement the work of other investigators presented in this volume. It should be emphasized that this work has not yet been published and is presented here as preliminary investigations. The thymus of course is the key organ for T cell development. It is essential for preventing autoimmunity by establishing tolerance as a result of imposing negative selection on autoreactive T cells.

First, we demonstrate that the expression of immunologically relevant genes varies across the panel of strains. This is illustrated for *Cd1d* in Fig 1. This gene encodes a member of the CD1 family that presents lipid antigens to NKT cells, a unique immunoregulatory T cell subset (10). In this example, the levels of *Cd1d* transcripts vary by a factor of at least two across the panel of strains.

Variation in expression levels of a particular gene is usually thought to be controlled by regulatory elements, such as promoters and enhancers, associated with that gene. It is becoming increasingly apparent, especially in complex genetic diseases, that gene-gene interactions are important in regulating expression levels. For example, polymorphisms in the *IL10* gene can affect the levels of IL-12 produced by dendritic cells (11). Quantitative trait loci (QTL; i.e. genes which have a quantitative effect on a particular trait) may be mapped using one of the WebQTL modules. In order to demonstrate this aspect of WebQTL, we focused on *Il7r*, the gene encoding the receptor for IL-7. This gene is of interest to those studying autoimmunity for a number of reasons, such as: the intrathymic level of IL-7 has a critical effect on the production of T cells; IL-7 signaling enhances the survival of developing thymocytes and mature T cells; IL-7 is critically regulated in developing thymocytes; thymocytes that undergo positive selection upregulate IL-7R expression; and IL-7 is a survival factor for CD4⁺ CD25⁺ T-cells (reviewed in refs 12, 13; 14).

An example of such QTL linkage mapping is presented in Fig 2. The *Il7r* gene itself is on chromosome 1, but there is no significant affect of polymorphisms in this region on variation in IL-7R expression levels. In contrast, a locus on chromosome 14 has the most significantly affect on thymic *Il7r* expression. Thus, the most powerful regulation of *Il7r* appears not to be by cis-elements associated with the *Il7r* promoter, but by a locus on another chromosome instead.

What would happen when *I17r* expression increases or decreases across the panel of strains? Do other transcripts also rise (or fall)? What genes may decrease in expression as *I17r* expression increases? In other words, can we correlate any other genes' expression with that of *I17r*? Using WebQTL to interrogate the data from 26 strains, it is possible to define a cluster of apparently co-regulated genes. Table 1 shows the results for the ten most highly correlated genes from such an analysis. The P values for the correlation of these genes are all highly significant ($P < 0.0000001$). The transcripts shown in Fig 3 are all positively correlated, but genes that have an inverse correlation can also be revealed by this method.

Perhaps the most powerful application of WebQTL comes from its ability to integrate the expression of each gene of a cluster with the underlying BXD strains' genotypes. In this way, we can map loci throughout the genome which may influence expression of each of these genes. The output of one such analysis is shown in Figure 3 in which each individual transcript of a cluster of 20 genes whose expression is correlated with that of *I17r* was examined. First, these transcripts are clustered in a tree according to their correlation status. Then variation in gene expression is integrated with the underlying genetic variation amongst the strains. This allows mapping of loci which can influence the expression of each of these genes, as depicted in the heat map: the more intense the colour, the stronger the effect of that locus. Even with this "helicopter view" one can see (from the black/grey areas) that most of the genome has no effect on expression of any of these genes; this is of course to be expected. What is also immediately noticeable is that there are several brightly coloured areas, which indicate the position of loci at which strains with similar expression levels share more of the parental B or D alleles than could be expected by chance. That is, these are loci showing linkage to these phenotypes; these loci contain genes which influence the levels of the transcripts in the *I17r* cluster.

Focusing on specific regions allows us to formulate hypotheses that may be tested subsequently. For example, Fig 3 shows the influence of a locus on chromosome 3 on the levels of all of these transcripts. Thus, the expression not just of *I17r* but also of each of the highly correlated genes is regulated by a master locus on chromosome 3. There is also another locus on chromosome 7 that influences expression of many of the genes; in particular it has its strongest effect on *Sh3bgr12* (SH3 binding glutamic acid-rich -2). Thus, a simple WebQTL session generates hypotheses that can initiate investigations into genes previously

unsuspected of having any functional involvement with the original reference gene, let alone with thymocyte development.

These examples give an indication of only some of the functions available within GeneNetwork. Currently, Systems Genetics is a powerful technology for defining clusters of co-regulated genes. Its use is centred upon user-specified genes and can identify novel potential master regulatory genes for further investigation. We are working to increase the functionality and power of the GeneNetwork and Systems Genetics further in a number of areas. In particular, increasing the number of strains studied can increase the mapping resolution. By increasing the genetic diversity of the founders of an RI set, the potential for observing regulatory polymorphisms increases dramatically. In this context, the availability of 1,000 RI strains from “The Gene Mine” (a.k.a. the Collaborative Cross) (15) will drive the development of Systems Genetics in further exciting areas.

Acknowledgements

This work was supported by grants from the University of Western Australia’s Strategy Fund, the Diabetes Research Foundation Ltd, and the National Health and Medical Research Council of Australia (#257601 and 305500), [to GM]; and by NIAAA U01AA13499 and P20-DA 21131 [to RWW].

References

1. Chesler E.J., Lu L., Wang J., Williams R.W., Manly K.F. 2004. WebQTL: rapid exploratory analysis of gene expression and genetic networks for brain and behavior. *Nature Neurosci.* 2004 7:485-6.
2. Chesler E.J., Lu L., Shou, S., Qu, Y., Gu, J., Wang J., Hsu, H.C., Mountz, J.D., Baldwin, N.E., Langston, M.A., Threadgill, D.W., Manly, KF & Williams, R.W. 2005. Complex trait analysis of gene expression reveals polygenic and pleiotropic networks that modulate nervous system function. *Nature Genetics.* 37:233-42.
3. Li H., Lu L., Manly K.F., Chesler E.J., Bao L., Wang J., Zhou M., Williams R.W., Cui Y. 2005. Inferring gene transcriptional modulatory relations: a genetical genomics approach. *Hum Mol Genet.* 14:1119-25.
4. Bystrykh L., Weersing E., Dontje B., Sutton S., Pletcher M.T., Wiltshire T., Su A.I., Vellenga E., Wang J., Manly K.F., Lu L., Chesler E.J., Alberts R., Jansen R.C., Williams R.W., Cooke M.P., de Haan G. 2005. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nature Genetics* 37:225-232
5. Hubner N., Wallace C.A., Zimdahl H., Petretto E., Schulz H., Maciver F., Mueller M., Hummel O., Monti J., Zidek V., Musilova A., Kren V., Causton H., Game L., Born G., Schmidt S., Muller A., Cook S.A, Kurtz T.W, Whittaker J., Pravenec M., Aitman T.J. 2005. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature Genetics.* Mar;37:243-53.
6. Williams R.B., Cotsapas C.J., Cowley M.J., Chan E., Nott D.J., Little P.F. 2006. Normalization procedures and detection of linkage signal in genetical-genomics experiments. *Nature Genetics* 38:855-6
7. Taylor B.A. 1989. Recombinant inbred strains. In: *Genetic variants and strains of the laboratory mouse* 2nd Ed Lyon ML, Searle AG, eds, pp 773-796. Oxford: Oxford UP.
8. Peirce J.L., Lu L., Gu J., Silver L.M., Williams R.W. 2004. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet.* 5:7.
9. Wang J., Williams R.W., Manly K.F. 2003. WebQTL: web-based complex trait analysis. *Neuroinformatics* 1:299-308.
10. Godfrey D.I., Berzins S.P. 2007. Control points in NKT-cell development. *Nat Rev Immunol.* 7:505-18.
11. Peng J.C., Abu Bakar S., Richardson M.M., Jonsson J.J., Frazer I.H., Nielsen L.K., Morahan G., Thomas R. 2006. *IL10* and *IL12B* polymorphisms each influence IL-12p70 secretion by dendritic cells in response to LPS. *Immunol Cell Biol.* 84:227-32.
12. Akashi K., Kondo M., Weissman I.L. 1998. Role of interleukin-7 in T-cell development from hematopoietic stem cells. *Immunol Rev.* 165:13-28.
13. Aspinall R. 2006 T cell development, ageing and Interleukin-7. *Mech Ageing Dev.* 127:572-8.
14. Harnaha J, Machen J, Wright M, Lakomy R, Styche A, Trucco M, Makaroun S, Giannoukakis N. 2006. Interleukin-7 is a survival factor for CD4+ CD25+ T-cells and is expressed by diabetes-suppressive dendritic cells. *Diabetes* 55:158-70.
15. The Complex Trait Consortium. 2004. The Collaborative Cross. *Nature Genetics.* 36:1133-7.

Figure Legends

Figure 1. The amount of *Cd1d* transcript in thymus from each of the strains was determined, and normalized by the rank-inverse method using Illumina Bead Studio software. Trait score is shown as a \log_2 value, and variation in expression amongst strains was plotted using WebQTL. Each strain is indicated by a '+'; the number of the BXD strain (7, 8) is shown immediately below this symbol. This plot can be used for assessing whether the strain data is normally distributed, as a normally distributed set should form a straight line.

Figure 2. Genome-wide linkage analysis of variation in *Il7r* expression amongst BXD strains. Evidence of linkage is presented as LOD scores (thick black lines) which are calculated at each physical position (indicated by megabase number) on every chromosome (indicated by numbers in each panel). Upper and lower horizontal dashed lines indicate genomewide significant and suggestive linkage thresholds, respectively. The location of the *Il7r* gene itself on chr. 15 is arrowed. Further information is displayed on the graph as thinner grey lines. A positive additive coefficient indicates that DBA/2J alleles increase trait values. In contrast, a negative additive coefficient indicates that C57BL/6J alleles increase trait values. NB: All this information is presented in colour by WebQTL and is more easily viewed than in this greyscale image.

Figure 3. Systems genetics analysis of the genes most highly correlated with *Il7r* in thymus. WebQTL integrated variation in expression across the 26 BXD strains of the 20 genes whose expression in the thymus was most highly correlated with that of *Il7r* (Table 1) with the underlying genetic variation of those strains. (See text). The figure depicts a heat map, with the genome arrayed on the Y axis, from chromosome 1 at the top through to the X chromosome; chromosomes are indicated by the numbers on the left of each block. The X axis contains all 20 genes to be tested: each "lane" represents a separate gene in the following order: *Il7r*, *Il18r1*, *B830021E24Rik*, *I200002N14Rik*, *Traf1*, *AA175286*, *Ctsw*, *Timp3*, *AA960558*, *Adam19*, *Icos*, *Ppp3cc*, *F2r*, *Sh3bgrl2*, *Ebi2*, *2610318I01Rik*, *D130072O21Rik*, *AU041483*, *Gpr18*, *Emp2*. Within each lane the likelihood that variation in expression of that gene is due to a polymorphism at each position is assessed and indicated by a colour. Here, the results are presented as a heat map, with individual effects of the parental C57BL/6 (B; blue) and DBA/2 (D; red) alleles on transcript levels indicated with different colours. The more intense the colour, the higher the likelihood that the allele contributes to the regulation. Information regarding the location of the gene encoding each of the transcripts is also provided, depicted by a triangle at the appropriate position within each lane.

Table 1. Genes whose variation in expression is correlated with that of *Il7r*.

Rank	Symbol	Chr	Megabase	Mean Expr	Correlation	p Value
1	<i>Il7r</i>	15	9.45	9.29	1.00	0.00E+00
2	<i>Il18r1</i>	1	40.45	9.15	0.89	1.55E-11
3	<i>1200002N14Rik</i>	10	87.75	8.91	0.85	1.91E-09
4	<i>Ppp3cc</i>	14	68.95	7.92	0.85	2.11E-09
5	<i>Ctsw</i>	19	5.47	9.25	0.84	7.06E-09
6	<i>Ebi2</i>	14	121.09	8.75	0.83	9.6E-09
7	<i>Timp3</i>	10	85.78	9.40	0.83	1.22E-08
8	<i>Gpr18</i>	14	121.05	9.14	0.83	1.25E-08
9	<i>B830021E24Rik</i>	16	44.16	8.72	0.82	2.10E-08
10	<i>Sh3bgrl2</i>	9	83.40	7.73	0.82	2.75E-08
11	<i>Icos</i>	1	60.94	8.61	0.82	4.10E-08
12	<i>Emp2</i>	16	10.20	7.89	0.81	4.91E-08
13	<i>2610318I01Rik</i>	3	122.53	9.26	0.81	5.88E-08
14	<i>AA960558</i>	7	130.70	10.83	0.81	9.15E-08
15	<i>AU041483</i>	16	36.65	6.85	0.80	1.04E-07
16	<i>Traf1</i>	2	34.77	9.18	0.80	1.09E-07
17	<i>Adam19</i>	11	45.99	8.44	0.80	1.21E-07
18	<i>F2r</i>	13	96.70	10.52	0.80	1.26E-07
19	<i>D130072O21Rik</i>	?	?	10.05	0.80	1.39E-07
20	<i>AA175286</i>	6	3.32	11.59	0.80	1.43E-07

Variation in gene expression of all transcripts on the Illumina Ref 8 chip was assessed. Values of expression of *Il7r* in the database were compared to all 24,607 records in Thymus Illumina RankInv(June07) database. The TOP 100 correlations ranked by the Pearson's product-moment correlation were calculated and the highest 20 correlated genes are displayed here. Also shown is the chromosomal position of that gene (chromosome number and megabase position), its mean expression levels (normalised, rank-inverted and expressed as log₂), the Pearson's product-moment correlation, and the P value corresponding to that correlation.





