Article

Combining Mouse Congenic Strains and Microarray Gene Expression Analyses to Study a Complex Trait: The NOD Model of Type 1 Diabetes

Iain A. Eaves,[1] Linda S. Wicker,[2,5] Ghassan Ghandour,[3] Paul A. Lyons,[1] Laurence B. Peterson,[4] John A. Todd,[1,6] and Richard J. Glynne[3]

[1]*Juvenile Diabetes Research Foundation/ Wellcome Trust (JDRF/WT) Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/Medical Research Council (MRC) Building, Addenbrooke's Hospital, Cambridge, CB2 2XY, UK;* [2]*Department of Immunology and Rheumatology, Merck Research Laboratories, Rahway, New Jersey 07065, USA;* [3]*Eos Biotechnology, Inc., South San Francisco, California 94080, USA;* [4]*Department of Pharmacology, Merck Research Laboratories, Rahway, New Jersey 07065, USA*

Combining congenic mapping with microarray expression profiling offers an opportunity to establish functional links between genotype and phenotype for complex traits such as type 1 diabetes (T1D). We used high-density oligonucleotide arrays to measure the relative expression levels of >39,000 genes and ESTs in the NOD mouse (a murine model of T1D and other autoimmune conditions), four NOD-derived diabetes-resistant congenic strains, and two nondiabetic control strains. We developed a simple, yet general, method for measuring differential expression that provides an objective assessment of significance and used it to identify >400 gene expression differences and eight new candidates for the *Idd9.1* locus. We also discovered a potential early biomarker for autoimmune hemolytic anemia that is based on different levels of erythrocyte-specific transcripts in the spleen. Overall, however, our results suggest that the dramatic disease protection conferred by six *Idd* loci (*Idd3*, *Idd5.1*, *Idd5.2*, *Idd9.1*, *Idd9.2*, and *Idd9.3*) cannot be rationalized in terms of global effects on the noninduced immune system. They also illustrate the degree to which regulatory systems appear to be robust to genetic variation. These observations have important implications for the design of future microarray-based studies in T1D and, more generally, for studies that aim to combine genome-wide expression profiling and congenic mapping.

[The supplemental research data accompanying this article are available through the authors' web site (http://www-gene.cimr.cam.ac.uk/todd/), and the array data have been submitted to the GEO data repository (http://www.ncbi.nlm.nih.gov/geo/) under accession no. GSE11]

The nonobese diabetic (NOD) mouse is the principle animal model of type 1 diabetes (T1D; Atkinson and Leiter 1999), an autoimmune disorder that results from the T cell-mediated destruction of the insulin-producing β cells in the pancreas (Pipeleers and Ling 1992). Like most common diseases with a substantial impact on public health, T1D is thought to result from a complex interplay of multiple genetic and environmental factors (Todd 1999). To date, genetic analysis has provided evidence for >20 murine diabetogenic loci (Lyons and Wicker 2000; Rogner et al. 2001), and at least this number probably exists in humans (Concannon et al. 1998; Mein et al. 1998). Whilst the identity of the genes involved may not be the same in the two species, it is expected that key physiological pathways involved in the pathogenesis of T1D will be conserved between mice and humans (Risch et al. 1993; Vyse and Todd 1996). Consequently, it is hoped that identification of the genes and pathways involved in murine T1D will shed light on the human condition.

Congenic mapping has been used to confirm the existence of and to fine map several murine T1D susceptibility loci (Lyons and Wicker 2000), and one locus, *Idd3*, has already been localized to an interval of <0.15 cM on mouse chromosome 3 (Lyons et al. 2000a). Nonetheless, outside the major genetic determinant, *Idd1*, which colocalizes with the major histocompatability class II genes on mouse chromosome 17, the identities of these putative *Idd* loci and the pathways in which they act remain unknown. Recent advances in technology mean that traditional genetic mapping techniques can now be complemented by high-throughput methods for studying gene function and regulation. High-density arrays of synthetic oligonucleotides (Lipshutz et al. 1999), or cDNAs (Schena et al. 1995), allow gene expression monitoring on a genome-wide scale and offer an opportunity to establish functional links between genotype and phenotype for complex diseases like T1D. Consequently, they are expected to aid in

[5]Present address: JDRF/WT Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Addenbrooke's Hospital, Cambridge, CB2 2XY, UK
[6]Corresponding author.
E-MAIL john.todd@cimr.cam.ac.uk; FAX 44-1223-762102.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.214102. Article published online before print in January 2002.

12:232–243 ©2002 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/02 $5.00; www.genome.org

the identification of novel susceptibility genes and biochemical pathways not previously known to be involved in disease etiology.

Microarray analysis has already been used to generate genome-wide expression profiles of certain yeast mutants (DeRisi et al. 1997; Zhu et al. 2000) and, in mice, it has helped characterize several transgenic lines at the molecular level (Callow et al. 2000; Aronow et al. 2001). It has also been used in combination with traditional quantitative trait locus (QTL) mapping techniques to successfully identify complement factor 5 (*C5*) as a susceptibility locus in a murine model of allergic asthma (Karp et al. 2000). A strategy that combines congenic mapping with microarray expression profiling promises to offer both these possibilities. Gene identification, a slow and laborious process, may be speeded up as in the case of *C5* and asthma. However, in addition, this approach offers the opportunity to explore the functional consequences of a defined (but not completely characterized) genetic difference at the molecular level before the identity of the disease susceptibility locus itself is known.

To date, only one study has combined a congenic strain strategy with microarray expression analysis (Aitman et al. 1999). Aitman and colleagues studied the spontaneously hypertensive (SHR) rat, a model for human insulin-resistance, type 2 diabetes, obesity, hyperlipidemia, and essential hypertension. Congenic mapping had placed a QTL affecting glucose and fatty acid metabolism in a 36-cM interval of rat chromosome 4, but the causative gene(s) had not been identified. cDNA microarrays were used to compare gene expression in adipose tissue from the control, SHR, and congenic strains. Three clones encoding rat *Cd36*, a gene known to map to regions of mouse and human chromosomes syntenic to rat chromosome 4, showed reduced hybridization signals for SHR compared with those from the control and congenic strains. Therefore, *Cd36* was pursued as a candidate for the QTL, and SHR was shown to have a functional deficiency of this fatty acid transporter and receptor in fat and heart. Sequencing revealed that the SHR *Cd36* cDNA contains multiple variants caused by unequal genomic recombination of a duplicated ancestral section. The apparently diminished expression of *Cd36* in SHR rats was traced to a genomic deletion within the 3′-UTR, the only region represented on the array.

Cd36 may represent an exceptional case, in which a physical deletion directly affected the expression level reported by the array. However, it seems likely that a similar approach will prove successful in identifying genes, the true expression of which is affected either directly, or indirectly, by defined genetic differences between parent and congenic strains. Here, we report the first application of polymorphic expression profiling to study T1D in the NOD mouse. We also present a simple, yet general framework for measuring differential gene expression that provides an objective assessment of significance rather than relying on ad hoc thresholds.

RESULTS

Experimental Outline

One of the first steps in designing a microarray expression study is identifying the most appropriate cell type or tissue to profile. This step is relatively straightforward when there is a clear target as in the case of cancer; however, it becomes less clear for other conditions such as inflammatory and autoimmune diseases. In these cases, multiple cell types may play a direct role in the pathologic process, and the action of a given susceptibility locus may be restricted to only one of them. Hence, a decision must be made about whether to cast the initial net wide, or restrict analysis to a well-defined population of cells. Here, we have adopted the former approach as the cell type(s) most relevant to the action of the different *Idd* loci are unknown. Owing to the importance of T cells in the development of T1D, we decided to target two organs, the spleen and the thymus. We profiled gene expression in thymi from 4-week-old female mice and spleens from 3-month-old female mice. Four weeks is a key time point in the development of diabetes in the NOD mouse, marking a point at which the infiltration of the pancreatic islets by mononuclear cells has begun (Dahlen et al. 1998). At 3 months, the mice are adult with fully developed immune systems, and the pathogenic process has been triggered in most NOD mice by this point. However, very few will have progressed to develop overt diabetes (Wicker et al. 1995). For each tissue, we profiled seven strains, NOD itself, four NOD-derived diabetes-resistant congenic strains, and two B10-derived nondiabetic control strains (Table 1; Fig. 1).

For each of the 14 strain–tissue combinations (7 strains $\times$ 2 tissues), we performed two independent replicate experiments, making the total number of hybridizations performed 28. In each case, an RNA population from a pool of two or three organs was generated to minimize the chances of within strain variation masking genuine variation between the strains. In addition, we processed samples within each replicate group in parallel to minimize experimental sources of variation (see Methods). To identify genes whose level of expression was influenced by a particular *Idd* locus/loci, we classified each of the five NOD strains according to the presence/absence of the NOD susceptibility allele(s) (Table 2). This scheme had two advantages over pairing each congenic strain separately with the NOD reference. First, the effective increase in sample size increased the power of our analysis, enabling us to detect much smaller changes in expression level than would otherwise have been possible. Second, it allowed us to assess the variability associated with a given observed change in gene expression. This feature was critical for establishing which genes were most likely to be genuinely differentially expressed, especially given the increase in variability associated with low intensity probe sets (Fig. 2). To control for the action of secondary nuisance loci in our analysis, for example, an effect of the *Idd9* locus in a comparison aimed at identifying genes for which expression was influenced by the *Idd3* locus, genes for which expression was affected by one locus were excluded from comparisons involving the other loci.

Measuring Differential Expression

We devised a modified *t* statistic to assess differential gene expression. In particular, we exploited the inherent parallel nature of microarray experiments to obtain more robust estimates of the within-group variances. That is, variances were estimated as a weighted average of the observed variance for a particular probe set and a mean local variance, estimated from probe sets with similar normalized hybridization intensities. This step acted to stabilize the variance estimates associated with individual genes, which were expected to be poor because of the small number of samples within each group (Fig. 3). Whilst this calculation provided a robust and objective measure of differential expression, it was not clear how much confidence should be placed in a given magnitude of *t* representing a genuine difference in gene expression. Previously,

Table 1. The Seven Mouse Strains Used in this Study

Strain	Abbreviated name	*Idd* loci in congenic segment(s)	Parental strain	Donor strain	Chromosome	Size of congenic segment (cM)	Diabetes incidence (females) (%)	Ref.
NOD	NOD	—	NOD	—	—	—	75–80	Harada and Makino 1992
NOD.B6 *Idd3* R450	*Idd3*	*Idd3*	NOD	C57BL/6	3	0.35	20–25	Denny et al. 1997
NOD.B10 *Idd5* R444	*Idd5*	*Idd5.1, Idd5.2*	NOD	C57BL/10	1	~10	35–45	Hill et al. 2000
NOD.B6 *Idd3* R450 B10 *Idd5* R8	*Idd3+5*	*Idd3, Idd5.1, Idd5.2*	NOD	C57BL/6, C57BL/10	3, 1	0.35, ~50	<2	Hill et al. 2000
NOD.B10 *Idd9* R28	*Idd9*	*Idd9.1, Idd9.2, Idd9.3*	NOD	C57BL/10	4	~45	3	Lyons et al. 2000b
B10.NOD $H2^{g7}$	B10.$H2^{g7}$	*Idd1*	C57BL/10	NOD	17	Not defined	0	Wicker et al. 1993
B10.NOD $H2^{g7}$ *Idd3*	B10.$H2^{g7}$ *Idd3*	*Idd1, Idd3*	C57BL/10	NOD	17, 3	Not defined, ~10	0	J.A. Todd and L.S. Wicker, unpubl.

A more detailed description of each congenic segment is given in Figure 1.

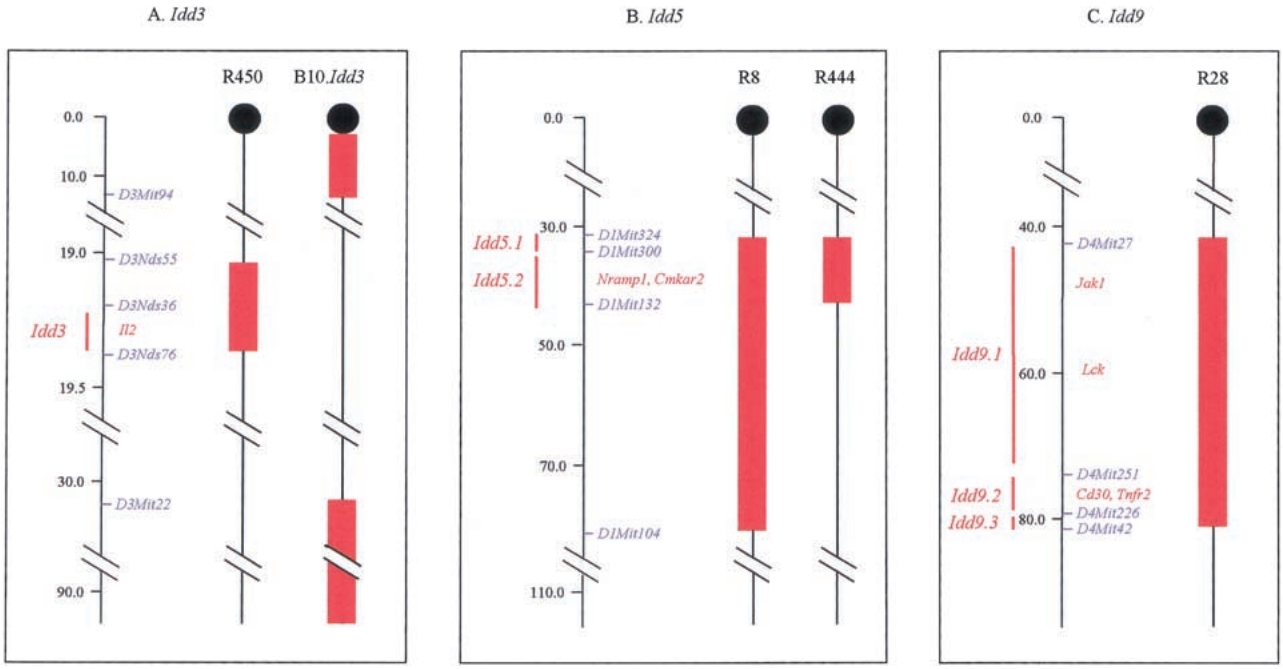

Figure 1 Genetic characterization of the congenic strains used. (Black lines) NOD genome; (red regions) location of the B10- or B6-derived congenic segments (with the exception of the B10.*Idd3* strain for which the congenic interval is NOD-derived and hence represented by a black line). (Blue text) Key microsatellites that define the boundaries of the congenic segments and/or the location of the *Idd* loci; (red vertical bars) position of each *Idd* locus. The locations of known candidate genes are shown. Because of space constraints, the candidates *Cflar*, *Casp8*, *Cd28*, and *Cd152* (*Ctla4*), which map to the *Idd5.1* locus, and *Cd137*, which maps to the *Idd9.3* locus, are not shown. All distances are given in centimorgans.

Callow et al. (2000) described a procedure whereby a permutation distribution of a *t* statistic could be generated through random assignment of each sample to the control and treatment groups. However, this method is restricted to highly replicated experiments and was not applicable here owing to the composition of the control group. Instead, we utilized the fact that we had carried out our entire experiment in duplicate to identify a control set of transcripts for each comparison. Those transcripts that appeared to be up-regulated in one experiment and down-regulated in the other were expected to have *t* values that were distributed independently of any genuine changes due to the congenic interval in question. We exploited this fact to set thresholds such that only a given number of genes were expected to exceed them by chance alone. An array-wide significance level of $P \leq 0.05$ (i.e., a type 1 error rate equivalent to 1 false positive for every 20 array-wide comparisons) was considered too stringent for the purposes of the present study, as it was expected to lead to the exclusion of genes that warranted further investigation. Consequently, thresholds were set such that only one of the ~20,000 transcripts deemed to be expressed was expected to exceed them by chance (Fig. 3). These changes were deemed suggestive. Gene expression differences that were likely to be the direct result of polymorphisms acting in *cis*, that is, primary changes, could then be identified through studying their chromosome locations and by comparing the expression levels in the relevant congenic and nondiabetic B10-derived strains. Finally, it is important to note that, as we have studied two complex tissues, any expression difference might reflect two distinct scenarios. First, it might be the result of a change in expression level in a particular constituent cell population(s). Second it might be due to a change in the relative

Table 2. Strain Groupings Used to Identify Differentially Expressed Genes

Idd locus/loci influencing expression level	Treatment group	Control group
Idd3	*Idd3, Idd3+5*	NOD, *Idd5, Idd9*
Idd5	*Idd5, Idd3+5*	NOD, *Idd3, Idd9*
Idd3+5	*Idd3+5*	NOD, *Idd3, Idd5, Idd9*
Idd9	*Idd9*	NOD, *Idd3, Idd5, Idd3+5*
Idd3, Idd5, Idd3+5, and *Idd9*	*Idd3, Idd5, Idd3+5, Idd9*	NOD
other	B10	NOD, *Idd3, Idd5, Idd3+5, Idd9*

For each comparison, the treatment group was defined as comprising those NOD-derived congenic strains carrying the C57BL/6, or B10 resistance allele at the locus of interest. Conversely, the control group consisted of those NOD-derived strains carrying the NOD susceptibility allele at the locus in question.

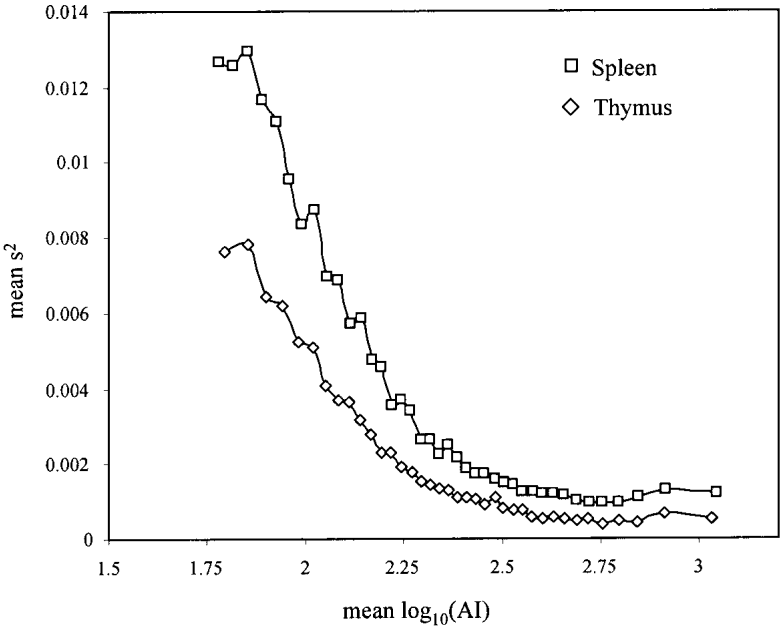

Figure 2 Relationship between mean $\log_{10}$(AI) and mean variance $\log_{10}$(AI) for the five NOD-derived strains (NOD, *Idd3*, *Idd5*, *Idd3/5*, *Idd9*). Probe sets were ordered according to their mean $\log_{10}$(AI) value across the five strains and assigned to successive, nonoverlapping bins of 500 probe sets. Variance across the five strains was calculated for each probe set independently, and the mean for each bin is plotted against the mean $\log_{10}$(AI) for that bin. As only a small proportion of genes were expected to be genuinely differentially expressed between the strains, the resulting pattern was anticipated to be representative of the underlying relationship between variance and $\log_{10}$(AI). The data shown are representative of both replicate experiments.

abundance of a certain cell population(s). In either case, the observed quantitative differences will reflect an underlying functional difference.

Summary of Gene Expression Changes

Idd3 and Idd5

Only eight probe sets were differentially expressed between strains carrying NOD- and B6-derived *Idd3* alleles (Table 3). The changes were modest (19%–66%), none of the genes were known to map to the 0.35-cM *Idd3* congenic interval, and collectively they did not implicate a pathway involved in *Idd3*-mediated disease protection/susceptibility (supplemental Table 1 [http://www-gene.cimr.cam.ac.uk/todd/]). Similarly, none of the genes whose expression was influenced by the *Idd5* congenic interval were known to map to the relevant ~40-cM segment of mouse chromosome 1 (supplemental Table 2A,B [http://www-gene.cimr.cam.ac.uk/todd/]). However, in the spleen, we observed a striking imbalance between the number of transcripts that appeared more/less abundant in congenic animals homozygous for the B10-derived *Idd5* interval (38 up-regulated vs. 1 down-regulated). This pattern was replicated in a comparison between the *Idd3+5* double congenic strain and its control group (30 up-regulated vs. 1 down-regulated; Table 3 and supplemental Table 3A [http://www-gene.cimr.cam.ac.uk/todd/]). In each case, this pattern could be explained, at least in part, by an increase in the level of multiple erythrocyte-specific mRNAs in the spleens of these strains (Fig. 4A). Interestingly, the majority of these transcripts were also more abundant in the NOD-related animals as a whole when compared with the B10.$H2^{g7}$ *Idd3* strain (the B10.$H2^{g7}$ strain was not available for comparison [see Methods]).

In the comparison involving the *Idd3+5* strain we identified two genes, chemokine (C-X-C) receptor 4 (*Cmkar4*) and serine/threonine kinase 25 (*Stk25*), that were expressed at a higher level in the thymus (four- to fivefold) as well as the spleen (two- to threefold). Moreover, gene expression was similarly elevated in the spleens and thymi of the B10.$H2^{g7}$ *Idd3* mice. Both *Cmkar4* and *Stk25* were known to map to part of mouse chromosome 1 contained within the *Idd5*R8 congenic interval but distal of the lower boundary of the introgressed segment in the single *Idd5*R444 strain (Fig. 1). As the expression profiles for these two genes correlated with the presence/absence of the NOD allele, we considered that *cis*-acting polymorphisms were the most likely explanation for these observations (Table 4).

A further 46 transcripts were identified as being present at a higher level in the thymi of *Idd3+5* mice, none of which were known to map to either of the congenic intervals in question (supplemental Table 3B [http://www-gene.cimr.cam.ac.uk/todd/]). Forty-four of these were also more abundant in the set 1 thymus sample of the B10.$H2^{g7}$ *Idd3* strain compared with the set 1 B10.$H2^{g7}$ sample. This pattern was not replicated among the second set of samples, suggesting that some contaminant neighboring tissue(s) could have been responsible for these additional differences (data not shown).

Idd9

In the spleens of *Idd9* mice, nine probe sets exhibited higher and nine probe sets exhibited lower hybridization intensities compared with the NOD control group (supplemental Table 4A [http://www-gene.cimr.cam.ac.uk/todd/]). Three of these probe sets were assigned to the same UniGene cluster, rhesus blood group-like (*Rhl1*), a gene known to map within the *Idd9*R28 congenic interval. All three registered lower expression levels for *Rhl1* in *Idd9* and B10.$H2^{g7}$ *Idd3* spleens relative to NOD. Four other genes were also known to map to this region of mouse chromosome 4 (Table 4). All four were expressed at lower levels in both the spleen and thymus of the *Idd9* strain, and all but one exhibited lower levels in the B10-derived strains. It is possible that the expression of this last gene, *lysosomal-associated protein 5* (*Laptm5*), is determined at least in part by the presence of specific NOD or B10 alleles at other loci explaining the observed difference in expression level between the *Idd9* and B10-derived strains. In the thymus, a total of 18 genes passed criteria that were modified following visual inspection of the data (supplemental Table 4B and supplemental Fig. 1 [http://www-gene.cimr.cam.ac.uk/todd/]). Of these, six were consistent with being the result of *cis*-acting polymorphisms, including three that had been previously identified in the spleen (Table 4).

Diabetes resistant/susceptible

All five of the NOD-derived congenic strains had been developed because they were, to varying degrees, resistant to de-

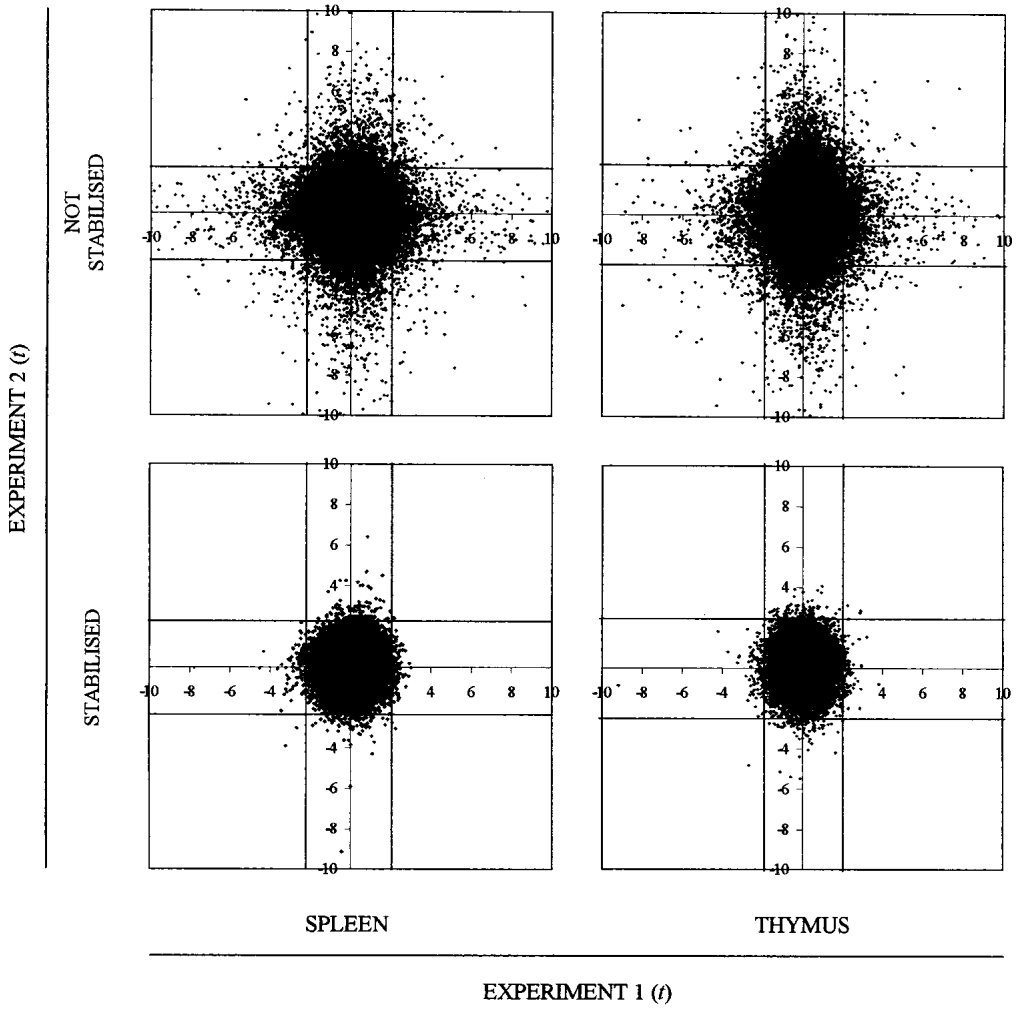

Figure 3 Joint distribution of *t* across the two independent replicate experiments for a comparison between NOD-derived strains carrying the *Idd3* resistance allele (*Idd3* and *Idd3/5*) and those carrying the susceptible allele (NOD, *Idd5*, and *Idd9*). (*Top* and *bottom*) Distributions before and after incorporation of the local variance estimates. Positive and negative values of *t* indicate stronger and weaker mean hybridization signals in the *Idd3–Idd3/5* group compared with the control group, respectively. The vertical and horizontal lines are set such that only 1 gene of the ~20,000 deemed to be expressed would be expected to fall in either the upper right or lower left quadrangles of the stabilized plots by chance. Eight probe sets met these criteria and were deemed to show suggestive evidence for a difference in expression level.

veloping diabetes in comparison with the parental NOD strain. To discover whether this shared difference had any common basis at the level of gene expression, we compared the expression profiles for these two sets of animals. A total of seven genes exhibited suggestive evidence for a difference in expression between the two groups, including the Thy1 antigen (CD90; supplemental Table 5 [http://www-gene.cimr.cam.ac.uk/todd/]).

B10 versus NOD

In addition to searching for genes that might be differentially expressed between the congenic strains and the parental NOD strain, we decided to compare mRNA levels of the NOD-like strains as a whole with those of the nondiabetic B10-like animals. This comparison gave us the opportunity to identify genes whose expression was affected by loci other than those represented in the panel of congenics. Unfortunately, we were only able to perform this comparison in the spleen as appropriate replicate samples were not available for the B10 thymus group (see Methods).

As expected, we identified more changes in gene expression (295) than in the other comparisons (supplemental Table 6 [http://www-gene.cimr.cam.ac.uk/todd/]), including two, *Ly6c* and *Fcgr2b*, that had previously been shown to be defectively expressed in the lymphocyte compartment of the NOD mouse (Philbrick et al. 1990; Luan et al. 1996; Pritchard et al. 2000). The most striking finding, however, was the increase in the level of a large number of red cell-specific transcripts in the spleens of the NOD-like strains (Fig. 4A). As discussed below, this observation may be explained by an increase in the number of erythrocyte precursors in the splenic red pulp of these mice. Surprisingly, we also observed differences in the levels of a number of transcripts characteristic of exocrine pancreatic tissue. In particular, expression of these genes was consistently higher in all the NOD-like strains relative to the B10.$H2^{g7}$ *Idd3* strain in both independent replicate sets (Fig. 4B).

Table 3. Summary of Suggestive Changes in Gene Expression Identified in Comparisons between the Strain Groups Shown in Table 2

Comparison		No. of changes (unique)		No. of potential 1° changes	No. of control changes
locus	tissue	up	down		
Idd3	S	1 (1)	2 (2)	0	2
	T	4 (3)	1 (1)	0	1
Idd5	S	38 (37)	1 (1)	0	1
	T	7 (7)	6 (6)	0	1
Idd3+5	S	30 (28)	1 (1)	2	2
	T	48 (48)	2 (2)	4	2
Idd9	S	9 (9)	9 (6)	7[a]	1
	T	7 (7)	4 (4)[b]	4[b]	2[b]
Idd3, Idd5, Idd3+5, Idd9	S	0 (0)	2 (2)	—	0
	T	2 (2)	3 (3)	—	1
other	S	141 (121)	153 (131)	—	0
total	—	287 (263)	186 (159)	17	13

Full details of all changes are given in the supplementary material.
[a]Only five genes were unique.
[b]Includes only genes that met the original criteria for a suggestive change. An additional six probe sets met the modified criteria for a suggestive decrease in expression, including two potential primary changes. A further five control probe sets also passed the modified criteria (see supplemental Fig. 1 [http: //www.gen.cimr.cam.ac.uk/todd/]).

Of the remaining 246 probe sets, 21 were known to represent genes that colocalized with published *Idd* loci other than those represented in our panel of congenic strains (Table 5). One of these, small inducible cytokine A5 (*Scya5* or RANTES), was of particular interest. *Scya5* was a strong functional candidate for the *Idd4* locus on mouse chromosome 11 (Todd et al. 1991; Gill et al. 1995; McAleer et al. 1995). Moreover, preliminary reports suggested that a variant that correlated with differential expression of the human gene was associated with a number of autoimmune conditions (Fryer et al. 2000; Makki et al. 2000; Nickel et al. 2000). Hence, we attempted to replicate our finding using an established, alternative platform. Analysis by RT-PCR TaqMan revealed that *Scya5* expression was ~1.5-fold higher in the spleens of B10-derived mice compared with NOD-derived strains, replicating the hybridization pattern observed with the oligonucleotide arrays (data not shown).

DISCUSSION

We combined a congenic strain strategy with microarray gene expression profiling to gain insight into the identity and action of six murine T1D susceptibility loci, *Idd3*, *Idd5.1*, *Idd5.2*, *Idd9.1*, *Idd9.2*, and *Idd9.3*. To this end, we compared the expression profiles of NOD mice, four NOD-derived congenic strains, and two nondiabetic control strains in two major T cell compartments, the spleen and thymus. Each congenic strain was homozygous for a B6/B10-derived resistance allele at one or more of the three *Idd* loci, and all were markedly protected from developing diabetes compared with the parent NOD strain (50% to >95% reduction).

We developed a simple, yet general framework for measuring differential gene expression that provides an objective assessment of significance rather than relying on ad hoc thresholds. Our approach was identical in spirit with the regularized *t*-test presented recently by Baldi and Long (2001) in that it exploits the inherently parallel nature of microarray studies to provide a more robust measure of differential expression in experiments with a low number of replicates. However, we evaluated significance empirically. We used independent replicate comparisons to identify a set of control transcripts that are expected to behave independently of the treatment being examined. This process allowed thresholds to be set such that only a given number of genes would be expected to exceed them by chance alone.

Using this approach, we identified a total of 472 suggestive changes in gene expression in 11 separate comparisons. This number compares with 13 control probe sets that passed the same thresholds, suggesting that the false-positive rate was very close to the desired value of one per array-wide comparison. Our analysis revealed two particularly unexpected findings. The first involved strain-specific differences in the level of certain erythrocyte-specific transcripts in the spleen. Levels were low (or absent) in B10.$H2^{g7}$ *Idd3* mice compared with NOD, *Idd3*, and *Idd9* mice, and higher levels still were observed for the two NOD congenic strains homozygous for the B10 diabetes resistance alleles at the *Idd5.1* and *Idd5.2* loci. It is known that aged (>200 days old) NOD mice can spontaneously develop Coombs'-positive hemolytic anemia (HA), a B cell-mediated organ-specific disease (Baxter and Mandel 1991). The reduced hematocrit (blood hemoglobin concentration) in HA represents a strong stimulus to murine splenic erythropoiesis (Pantel et al. 1990) and could account for the increase in the level of erythrocyte-specific transcripts observed in the spleens of the NOD-like strains. Previously, Baxter and Mandel (1991) were not able to detect signs of HA using the Coombs test in NOD mice as young as those studied here (~90 days old). However, it is possible that we have been able to detect the effects of hemolysis at a much earlier stage by profiling gene expression in the spleens of these animals. These results could lead to the development of a biomarker for the prediction of HA. Alternatively, our data may reflect the accumulation of post-transcriptional erythrocytes in the spleens of NOD mice. Regardless, our results suggest that a locus controlling this phenotype maps within the *Idd5*R444

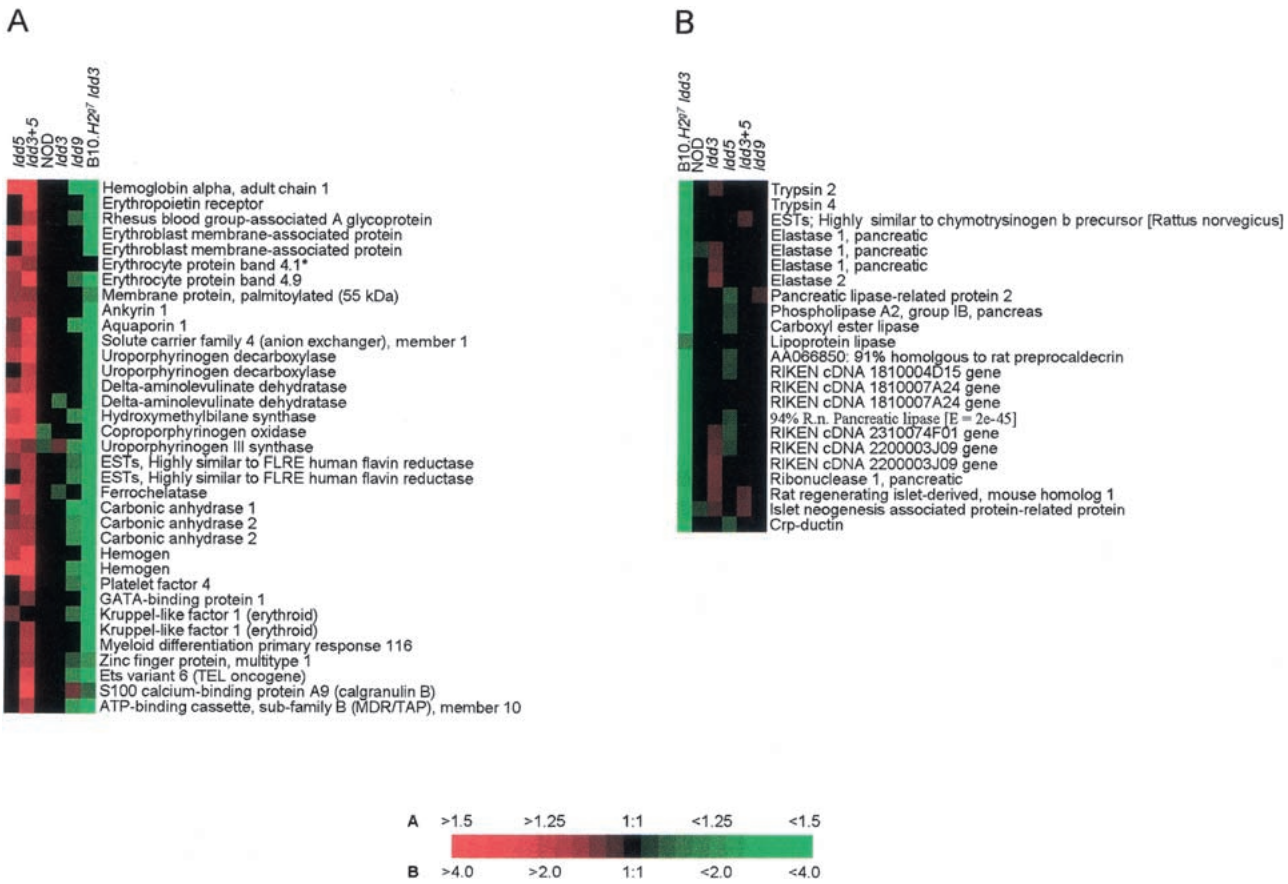

Figure 4 Relative expression across six strains for (*A*) all erythrocyte-specific genes and (*B*) all pancreas-specific genes exhibiting a suggestive difference in expression in at least 1 of the 11 comparisons performed. Expression levels were averaged across the two replicate experiments and converted to false color by use of the software TreeView (http://rana.lbl.gov/EisenSoftware.htm). (Red) Increase and (green) decrease in expression level relative to the median value for the six strains. Full details of all expression differences are given in the supplementary material.

congenic interval on mouse chromosome 1. Jordan et al. (2000) recently reported two genomic regions linked to *Mycobacterium bovis* induced HA in a NOD/BALB/c cross, one on chromosome 17 (*Bah1*) and one on chromosome 16 (*Bah2*). They also reported suggestive evidence for linkage on chromosome 1; however, this region was 20–40 cM distal of the lower boundary of the *Idd5*R444 interval. Moreover, it was the NOD allele that contributed susceptibility in this case.

The second unexpected example of a cluster of functionally related genes with lower expression in the spleens of B10 mice is a group encoding enzymes characteristic of the exocrine pancreas. Although the pancreas is anatomically close to the spleen it is highly unlikely that the NOD-like samples could have been contaminated with pancreatic tissue. Moreover, random carryover of pancreatic material would be strain independent and result in different preparations being con-

Table 4. Gene Expression Changes that Were Consistent with Being a Direct Result of cis Acting Polymorphisms

Gene	UniGene ID	Strain	*Idd* locus	Idd (up/down)	Tissues observed	B10 level	Maps to interval
Chemokine (C–X–C) receptor 4	Mm.1401	*Idd3+5*	—	↑	S, T	*Idd3+5*	yes
Serine/threonine kinase 25	Mm.28761	*Idd3+5*	—	↑	S, T	*Idd3+5*	yes
ESTs, moderately similar to ladinin [*Homo sapiens*]	Mm.36726	*Idd3+5*	—	↑	T	*Idd3+5*	unknown
cDNA clone IMAGE:1277236	—	*Idd3+5*	—	↑	T	*Idd3+5*	unknown
Rhesus blood group-like	Mm.18140	*Idd9*	*Idd9.1*	↓	S	*Idd9*	yes
High mobility group nucleosomal binding domain 2	Mm.911	*Idd9*	*Idd9.1*	↓	S, T	*Idd9*	yes
RIKEN cDNA 9030402K04 gene	Mm.43213	*Idd9*	*Idd9.1*	↓	S	*Idd9*	yes
RIKEN cDNA 3200001F09 gene	Mm.88349	*Idd9*	*Idd9.1*	↓	S, T	*Idd9*	yes
Lysosomal-associated protein transmembrane 5	Mm.4554	*Idd9*	*Idd9.1*	↓	S, T	*NOD*	yes
Eph receptor A2	Mm.2581	*Idd9*	*Idd9.1*	↓	T	*Idd9*	yes
Transcription elongation factor B (SIII)	Mm.27663	*Idd9*	*Idd9.1*	↑	T	*Idd9*	yes
RIKEN cDNA 4930506L07 gene	Mm.22176	*Idd9*	*Idd9.1*	↑	T	*Idd9*	yes

Table 5. Identities of the 21 Probes Sets Representing Genes that Were Differentially Expressed between the NOD and B10 Strain Groups and Were Known to Co-localize with Published *Idd* Loci Other than *Idd3*, *Idd5*, and *Idd9*

Idd locus	Gene	UniGene	Fold-change (B10/NOD)
Idd1	Mouse MHC class I Q8/9d cell surface antigen	Mm.88752	1.30
Idd1	Mouse MHC class I Q8/9d cell surface antigen	Mm.88752	1.28
Idd1	MHC [A.CA/J(H-2K-f)] class I antigen	Mm.34332	1.19
Idd1	Histocompatibility 2, class II, locus DMa	Mm.16373	1.22
Idd1	Histocompatibility 2, class II antigen E beta	Mm.22564	1.29
Idd1	Histocompatibility 2, class II antigen E beta	Mm.22564	1.23
Idd1	Histocompatibility 2, K region	Mm.16771	1.33
Idd1	Histocompatibility 2, M region locus 3	Mm.14437	1.30
Idd1	Histocompatibility 2, T region locus 10	Mm.87776	1.41
Idd1	Histocompatibility 2, T region locus 22	Mm.88783	1.31
Idd1	Histocompatibility 2, T region locus 23	Mm.35016	1.25
Idd4	Small inducible cytokine A5 (Scya5)	Mm.3370	1.56
Idd6	Kirsten rat sarcoma oncogene 2, expressed	Mm.31530	7.41
Idd7	CD79a antigen	Mm.1355	1.31
Idd7	CEA-related cell adhesion molecule 2	Mm.117098	1.61
Idd7	Paired-Ig-like receptor A1	Mm.87798	−1.43
Idd7	Avian reticuloendotheliosis viral (v-*rel*) oncogene related B	Mm.1741	1.67
Idd7	99% Avian reticuloendotheliosis viral (v-*rel*) oncogene related B	Mm.1741	1.37
Idd8	Vinculin	Mm.27206	−1.26
Idd8/12	Protein kinase C, delta	Mm.2314	1.29
Idd13	Proliferating cell nuclear antigen	Mm.7141	1.24

taminated to different degrees. In contrast, we observed similar, high transcript levels in all five NOD-derived animals across two completely independent replicate experiments. It is interesting to speculate that certain islet-specific transcripts might also be expressed in the NOD spleen where they might be involved in priming an antigen-specific immune response. However, as exocrine pancreatic enzymes are expressed at levels that are higher than those for almost any other genes in any organ (R.J. Glynne, unpubl.), a more sensitive technology would likely be required to detect signature islet-specific transcripts in the spleens of NOD mice.

Our decision to profile two complex tissues meant that any measurements necessarily represented the average of many different cell types. Consequently, we may have missed meaningful gene expression differences that were specific to certain cell populations, or masked by natural biological variation unrelated to the pathogenic process. Nonetheless, we detected the previously reported differences in the expression of *Ly6c* (Philbrick et al. 1990) and *Fcgr2b* (Luan et al. 1996; Pritchard et al. 2000) between NOD and B10-derived strains. We also observed a difference in the expression of the chemokine *Scya5* (RANTES) between these two strain groups. *Scya5* maps to the *Idd4* locus and is a persuasive candidate gene as it is known to act on T cells with a number of important consequences, including costimulation of cytokine release and T cell proliferation (Ward and Westwick 1998). *Idd4* colocalizes with *eae7* (Butterfield et al. 1998), a QTL in the principal animal model of multiple sclerosis, experimental allergic encephalomyelitis (EAE). The cDNA of *Scya5* has previously been screened for polymorphisms as part of a study to identify the etiological variant(s) at *eae7* (Teuscher et al. 1999). No coding polymorphisms were detected between B10.S/DvTe and SJL/J mice, however, a potentially polymorphic poly[d(C–A)] stretch is known to exist at position −800 to −763 (Danoff et al. 1994). A similar stretch is found 5′ of the structural gene for the chemokine MIP-1α and has been postulated to have enhancer-like activity (Widmer et al. 1991). Preliminary data suggest that a functional polymorphism in the promoter of the human gene is associated with atopic dermatitis (Nickel et al. 2000), asthma (Fryer et al. 2000), polymyalgia rheumatica, and rheumatoid arthritis (Makki et al. 2000). In each case, the susceptibility allele is associated with significantly higher constitutive expression of *Scya5* in vitro (Nickel et al. 2000). This finding is consistent with the increased expression of *Scya5* seen in a wide range of inflammatory disorders and pathologies (Appay and Rowland-Jones 2001; Gerard and Barrett 2001), including pancreatic infiltrates that promote rapid destruction of the insulin producing β-cells in the NOD mouse (Bradley et al. 1999). Interestingly, our results suggest that *Scya5* mRNA levels are actually lower in the spleens of NOD mice compared with the nondiabetic B10 strain. Further work will be required to show whether this difference is caused by allelic variation within the gene itself and, if so, whether this is the basis for the *Idd4* locus.

Whilst the differential expression of *Scya5* represents an interesting lead in the hunt for *Idd4*, our results suggest that more targeted experiments will be required to identify the genes and pathways involved in *Idd3*-, *Idd5*-, and *Idd9*-mediated T1D susceptibility/protection. We identified a striking increase in the expression of the functional candidate gene *Cmkar4* in *Idd3+5* and B10-like mice in both the spleen and thymus, consistent with a functional polymorphism(s) in *Cmkar4* itself. However, the gene has previously been excluded as a candidate for the *Idd5.1* and *Idd5.2* loci (Hill et al. 2000). In contrast, the true disease variants appear to affect gene expression in a more subtle, or at least transitory way. For example, despite conferring a 70% reduction in disease risk, only seven genes met the criteria for a suggestive difference in expression between strains carrying the NOD *Idd3* susceptibility allele and those carrying the B6 resistance allele. There were no obvious functional candidates amongst these

genes and no evidence for a particular pathway being disturbed. We did, however, identify eight genes in the *Idd9* interval that had expression patterns consistent with the existence of strain-specific allelic variation within the genes themselves. Although none represent obvious T1D susceptibility genes, they must still be considered new candidates for the *Idd9.1* locus.

Overall, our results reveal that the dramatic disease protection conferred by each of the three *Idd* loci cannot be rationalized in terms of global effects on the non-induced immune system. They also illustrate the degree to which regulatory systems appear to be robust to genetic variation. These observations have important implications for the design of future microarray-based studies in T1D and, more generally, for studies that aim to combine genome-wide expression profiling and congenic mapping. The existence of relatively few differences in gene expression between strains, even when a large chromosome segment is derived from two genetically distant strains, suggests that researchers will not be left searching through a sea of noise to identify those changes that are relevant to the phenotype in question. Therefore, combining congenic strains and microarray expression analysis is expected to be a powerful and specific approach for establishing functional links between genotype and phenotype for complex traits. Of course, the success of this approach will depend on choosing the correct target. Our results for T1D indicate the need to extend the present analysis to the activated/induced immune system. In addition, individual cell populations may have to be studied in a context-specific manner, for example, dendritic cells from different lineages at different stages of maturation. This approach will require the use of less expensive platforms such as medium-density spotted oligonucleotide/cDNA arrays. Interval-specific and immunospecific arrays can be printed at a fraction of the cost of commercially available high-density probe arrays, massively increasing the range of hypotheses that can be tested. We anticipate that, in combination with the congenic strains and analytical methods used here, these tools will enable us to make further unexpected discoveries that will shed light on the biology of the NOD mouse and, ultimately, relate specific observations to allelic variation within particular genes.

METHODS

Probe Arrays

We used custom Affymetrix GeneChips designed by Eos Biotechnology, Inc. The Eos custom GeneChips are designed to measure the expression of a larger number of genes or ESTs than the commercially available arrays. This goal is achieved by choosing those probes from within each set of 20 Affymetrix perfect match probes that show the least random fluctuation relative to the perceived specific hybridization over a wide ranging set of samples (Glynne et al. 2000). Experiments undertaken by Eos have revealed that the mismatch probes do not increase sensitivity or reproducibility (R.J. Glynne and G. Ghandour unpubl.). Consequently, each gene on the Eos array is represented by 6–8 perfect match probes, as opposed to 20 perfect match and 20 mismatch probes, increasing the number of genes for which expression can be assayed in parallel to >39,000.

Target Preparation and Microarray Hybridization

For each of the 28 independent samples, total RNA was extracted using the Trizol reagent from a pool of two or three thymi/spleens taken from age-matched female mice kept in specific pathogen-free conditions. To minimize any variation introduced during target preparation, samples were divided into two groups that would form independent replicate experiments. All samples within each group were processed in parallel. Three samples from group two (spleen: B10.$H2^{g7}$; thymus: B10.$H2^{g7}$, B10.$H2^{g7}$ *Idd3*) and one from group one (spleen B10.$H2^{g7}$) failed a preliminary quality control check. Consequently, the steps involved in preparing labeled cRNA from the initial starting material had to be repeated for these four samples at a later date. A subsequent comparison of the variation seen between samples within the same replicate group (samples processed in parallel) versus samples in distinct replicate groups (samples processed at different times) revealed that the variation between the two groups (i.e., owing to independent sample preparation) was higher than the small amount of variation between the closely matched strains (data not shown). Therefore, ignoring the variability introduced during sample preparation was likely to result in spurious changes being detected and genuine strain-specific differences in gene expression being masked. Consequently, we decided to exclude the four samples that had had to be prepared independently owing to failing initial quality control.

Target cRNA was prepared and hybridized to the Eos GeneChip arrays as described for standard, commercially available Affymetrix GeneChips (Mahadevappa and Warrington 1999), and raw image data was analyzed using the GeneChip Expression Analysis Software (Affymetrix). Data for each GeneChip was normalized using a proprietary method developed at Eos (Ghandour and Glynne 2000). Briefly, for each probe array background subtracted average cell intensities were fitted to a gamma distribution. These normalized cell intensities were then used to calculate an average intensity (AI) for each probe set. The AI was calculated as the trimean (T) of the probes making up a given probe set (Tukey 1977). These AI values were subjected to a second round of normalization by setting the 70th and 90th percentiles equal to the same value for each array.

Annotation

Accession numbers were used to search build #84 of the mouse UniGene dataset (http://www.ncbi.nlm.nih.gov/UniGene/). Each UniGene cluster was already extensively annotated and included mapping data sourced from the Mouse Genome Database (MGD) (http://www.informatics.jax.org/) and UniSTS (http://www.ncbi.nlm.nih.gov/genome/sts/). Sequences that were not represented in any UniGene cluster were compared with the GenBank non-redundant (nr) database (Benson et al. 2000) by BLAST (Altschul et al. 1990), and, where appropriate, homologies are reported.

Statistical Analysis

Some probe sets had negative AI values post-normalization and were set to a new minimum of one AI unit before log transformation. Probe sets were excluded from the analysis if the average mean AI values for the two strain groups in question fell below 50, as this value was not considered to be significantly above background. Differences in gene expression between two strain groups were assessed by calculating a modified t statistic. When at least two measurements were available for each group, t was calculated as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left\{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right\}}},$$

where $\bar{x}_1$ is the mean of the log AI values for the strains in the first group, $\bar{x}_2$ is the mean of the log AI values for the strains in the second group, n_1 is the number of strains in

group 1, n_2 is the number of strains in group 2 and s_1^2,s_2^2 are the pooled variance estimates for each group. Pooled variances were estimated as a weighted average of the actual observed variance and a mean local variance estimate, using a 2 : 1 ratio in favor of the larger of the two variance estimates. Weighting in favor of the larger of the two variance estimates provided a filter to screen out real changes that were not due to the presence/absence of the congenic interval in question, for example, changes due to carryover of small amounts of neighboring tissue.

To calculate the mean local variance, probe sets in each group were ordered according to their mean log AI. Variances were then estimated for each of the 250 probe sets immediately above and each of the 250 below the probe set of interest. For each group, the average of these 500 values was then taken as the local variance estimate. Where one group was represented by a single sample, an analogous expression for t was used that only incorporated the variance estimates for the other group:

$$t = \frac{x_1 - \bar{x}_2}{\sqrt{\left\{s_2^2\left(1 + \frac{1}{n_2}\right)\right\}}}$$

A list of gene expression changes most likely to be genuine was generated by applying a filter that required a minimum threshold value for t in each replicate. These thresholds were estimated empirically for each comparison. Specifically, genes with positive values for t in one replicate group and negative values in the other (and vice versa) could be regarded as intrinsic controls, that is, transcripts that appeared to be up-regulated in one set but down-regulated in the other could be used to establish the distribution of t values for each replicate set independently of any real changes due to the congenic interval in question. Threshold values could then be calculated under the assumption that each probe set behaved independently. This assumption was clearly a conservative one; however, it greatly simplified estimation of the value of t required for a given expected number of false positives. As each comparison consisted of two experiments, we required that the value of t in each comparison exceeded the 99.5th percentile for the control distribution. The estimated false-positive rate per comparison is then $(1 - 0.995)$ [error rate expt. 1] $\times$ $(1 - 0.995)$ [error rate expt. 2] $\times$ 20,000 [number of trials] $\times$ 2 [each direction] = 1. A small degree of contamination of the spleen and thymus tissues with other material was conceivable and was likely to artificially inflate the threshold values. Therefore, the scatter plot for each comparison was visually assessed for evidence of an unusual number of outliers that were not consistent across the two replicates. This assessment led to the thresholds for one comparison being altered.

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust and the Juvenile Diabetes Research Foundation International. We thank Ed Tom, Dorian Willhite, and Jerry Lee for their help and contribution to this work.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Aitman, T.J., Glazier, A.M., Wallace, C.A., Cooper, L.D., Norsworthy, P.J., Wahid, F.N., Al-Majali, K.M., Trembling, P.M., Mann, C.J., Shoulders, C.C., et al. 1999. Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet.* **21:** 76–83.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Appay, V. and Rowland-Jones, S.L. 2001. RANTES: A versatile and controversial chemokine. *Trends Immunol.* **22:** 83–87.

Aronow, B.J., Toyokawa, T., Canning, A., Haghighi, K., Delling, U., Kranias, E., Molkentin, J.D., and Dorn, G.W. 2001. Divergent transcriptional responses to independent genetic causes of cardiac hypertrophy. *Physiol. Genomics* **6:** 19–28.

Atkinson, M.A. and Leiter, E.H. 1999. The NOD mouse model of type 1 diabetes: As good as it gets? *Nature Med.* **5:** 601–604.

Baldi, P. and Long, A.D. 2001. A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes. *Bioinformatics* **17:** 509–519.

Baxter, A.G. and Mandel, T.E. 1991. Hemolytic anemia in non-obese diabetic mice. *Eur. J. Immunol.* **21:** 2051–2055.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., and Wheeler, D.L. 2000. GenBank. *Nucleic Acids Res.* **28:** 15–18.

Bradley, L.M., Asensio, V.C., Schioetz, L.K., Harbertson, J., Krahl, T., Patstone, G., Woolf, N., Campbell, I.L., and Sarvetnick, N. 1999. Islet-specific Th1, but not Th2, cells secrete multiple chemokines and promote rapid induction of autoimmune diabetes. *J. Immunol.* **162:** 2511–2520.

Butterfield, R.J., Sudweeks, J.D., Blankenhorn, E.P., Korngold, R., Marini, J.C., Todd, J.A., Roper, R.J., and Teuscher, C. 1998. New genetic loci that control susceptibility and symptoms of experimental allergic encephalomyelitis in inbred mice. *J. Immunol.* **161:** 1860–1867.

Callow, M.J., Dudoit, S., Gong, E.L., Speed, T.P., and Rubin, E.M. 2000. Microarray expression profiling identifies genes with altered expression in HDL-deficient mice. *Genome Res.* **10:** 2022–2029.

Concannon, P., Gogolin-Ewens, K.J., Hinds, D.A., Wapelhorst, B., Morrison, V.A., Stirling, B., Mitra, M., Farmer, J., Williams, S.R., Cox, N.J., et al. 1998. A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. *Nature Genet.* **19:** 292–296.

Dahlen, E., Dawe, K., Ohlsson, L., and Hedlund, G. 1998. Dendritic cells and macrophages are the first and major producers of TNF-alpha in pancreatic islets in the nonobese diabetic mouse. *J. Immunol.* **160:** 3585–3593.

Danoff, T.M., Lalley, P.A., Chang, Y.S., Heeger, P.S., and Neilson, E.G. 1994. Cloning, genomic organization, and chromosomal localization of the Scya5 gene encoding the murine chemokine RANTES. *J. Immunol.* **152:** 1182–1189.

Denny, P., Lord, C.J., Hill, N.J., Goy, J.V., Levy, E.R., Podolin, P.L., Peterson, L.B., Wicker, L.S., Todd, J.A., and Lyons, P.A. 1997. Mapping of the IDDM locus Idd3 to a 0.35-cM interval containing the interleukin-2 gene. *Diabetes* **46:** 695–700.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278:** 680–686.

Fryer, A.A., Spiteri, M.A., Bianco, A., Hepple, M., Jones, P.W., Strange, R.C., Makki, R., Tavernier, G., Smilie, F.I., Custovic, A., et al. 2000. The -403 G→A promoter polymorphism in the RANTES gene is associated with atopy and asthma. *Genes Immunity* **1:** 509–514.

Gerard, C. and Barrett, J.R. 2001. Chemokines and disease. *Nature Immunol.* **2:** 108–115.

Ghandour, G. and Glynne, R. 2000. Method and apparatus for analysis of data from biomolecular arrays. International patent: WO0079465.

Gill, B.M., Jaramillo, A., Ma, L., Laupland, K.B., and Delovitch, T.L. 1995. Genetic linkage of thymic T-cell proliferative unresponsiveness to mouse chromosome 11 in NOD mice. A possible role for chemokine genes. *Diabetes* **44:** 614–619.

Glynne, R.J., Ghandour, G., and Goodnow, C.C. 2000. Genomic-scale gene expression analysis of lymphocyte growth, tolerance and malignancy. *Curr. Opin. Immunol.* **12:** 210–214.

Harada, M. and Makino, S. 1992. Biology of the NOD mouse. *Annual Report of the Shionogi Research Laboratories* **42:** 70–99.

Hill, N.J., Lyons, P.A., Armitage, N., Todd, J.A., Wicker, L.S., and Peterson, L.B. 2000. NOD Idd5 locus controls insulitis and diabetes and overlaps the orthologous CTLA4/IDDM12 and NRAMP1 loci in humans. *Diabetes* **49:** 1744–1747.

Jordan, M.A., Silveira, P.A., Shepherd, D.P., Chu, C., Kinder, S.J., Chen, J., Palmisano, L.J., Poulton, L.D., and Baxter, A.G. 2000. Linkage analysis of systemic lupus erythematosus induced in diabetes-prone nonobese diabetic mice by *Mycobacterium bovis*. *J. Immunol.* **165:** 1673–1684.

Karp, C.L., Grupe, A., Schadt, E., Ewart, S.L., Keane-Moore, M., Cuomo, P.J., Kohl, J., Wahl, L., Kuperman, D., Germer, S., et al. 2000. Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nature Immunol.* **1:** 221–226.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. *Nature Genet.* **21:** 20–24.

Luan, J.J., Monteiro, R.C., Sautes, C., Fluteau, G., Eloy, L., Fridman, W.H., Bach, J.F., and Garchon, H.J. 1996. Defective Fc gamma RII gene expression in macrophages of NOD mice: Genetic linkage with up-regulation of IgG1 and IgG2b in serum. *J. Immunol.* **157:** 4707–4716.

Lyons, P.A., Armitage, N., Argentina, F., Denny, P., Hill, N.J., Lord, C.J., Wilusz, M.B., Peterson, L.B., Wicker, L.S., and Todd, J.A. 2000a. Congenic mapping of the type 1 diabetes locus, Idd3, to a 780-kb region of mouse chromosome 3: Identification of a candidate segment of ancestral DNA by haplotype mapping. *Genome Res.* **10:** 446–453.

Lyons, P.A. and Wicker, L.S. 2000. Localising quantitative trait loci in the NOD mouse model of type 1 diabetes. In *Genes and Genetics in autoimmunity* (ed. A. Theofilopoulos), pp. 208–225. Harger, Basel, Switzerland.

Lyons, P.A., Hancock, W.W., Denny, P., Lord, C.J., Hill, N.J., Armitage, N., Siegmund, T., Todd, J.A., Phillips, M.S., Hess, J.F., et al. 2000b. The NOD Idd9 genetic interval influences the pathogenicity of insulitis and contains molecular variants of Cd30, Tnfr2, and Cd137. *Immunity* **13:** 107–115.

Mahadevappa, M., and Warrington, J.A. 1999. A high-density probe array sample preparation method using 10- to 100-fold fewer cells. *Nature Biotechnol.* **17:** 1134–1136.

Makki, R.F., al Sharif, F., Gonzalez-Gay, M.A., Garcia-Porrua, C., Ollier, W.E., and Hajeer, A.H. 2000. RANTES gene polymorphism in polymyalgia rheumatica, giant cell arteritis and rheumatoid arthritis. *Clin. Exp. Rheumatol.* **18:** 391–393.

McAleer, M.A., Reifsnyder, P., Palmer, S.M., Prochazka, M., Love, J.M., Copeman, J.B., Powell, E.E., Rodrigues, N.R., Prins, J.-P., Serreze, D.V., et al. 1995. Crosses of NOD mice with the related NON strain: A polygenic model for type 1 diabetes. *Diabetes* **44:** 1186–1195.

Mein, C.A., Esposito, L., Dunn, M.G., Johnson, G.C., Timms, A.E., Goy, J.V., Smith, A.N., Sebag-Montefiore, L., Merriman, M.E., Wilson, A.J., et al. 1998. A search for type 1 diabetes susceptibility genes in families from the United Kingdom. *Nature Genet.* **19:** 297–300.

Nickel, R.G., Casolaro, V., Wahn, U., Beyer, K., Barnes, K.C., Plunkett, B.S., Freidhoff, L.R., Sengler, C., Plitt, J.R., Schleimer, R.P., et al. 2000. Atopic dermatitis is associated with a functional mutation in the promoter of the C-C chemokine RANTES. *J. Immunol.* **164:** 1612–1616.

Pantel, K., Loeffler, M., Bungart, B., and Wichmann, H.E. 1990. A mathematical model of erythropoiesis in mice and rats. Part 4: Differences between bone marrow and spleen. *Cell Tissue Kinet.* **23:** 283–297.

Philbrick, W.M., Maher, S.E., Bridgett, M.M., and Bothwell, A.L. 1990. A recombination event in the 5′ flanking region of the Ly-6C gene correlates with impaired expression in the NOD, NZB and ST strains of mice. *EMBO J.* **9:** 2485–2492.

Pipeleers, D. and Ling, Z. 1992. Pancreatic beta cells in insulin-dependent diabetes. *Diabetes Metabol. Rev.* **8:** 209–227.

Pritchard, N.R., Cutler, A.J., Uribe, S., Chadban, S.J., Morley, B.J., and Smith, K.G. 2000. Autoimmune-prone mice share a promoter haplotype associated with reduced expression and function of the Fc receptor FcgammaRII. *Curr. Biol.* **10:** 227–230.

Risch, N., Ghosh, S., and Todd, J.A. 1993. Statistical evaluation of multiple-locus linkage data in experimental species and its relevance to human studies: Application to nonobese diabetic (NOD) mouse and human insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **53:** 702–714.

Rogner, U.C., Boitard, C., Morin, J., Melanitou, E., and Avner, P. 2001. Three loci on mouse chromosome 6 influence onset and final incidence of type I diabetes in NOD.C3H congenic strains. *Genomics* **74:** 163–171.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:** 467–470.

Teuscher, C., Butterfield, R.J., Ma, R.Z., Zachary, J.F., Doerge, R.W., and Blankenhorn, E.P. 1999. Sequence polymorphisms in the chemokines Scya1 (TCA-3), Scya2 (monocyte chemoattractant protein (MCP)-1), and Scya12 (MCP-5) are candidates for eae7, a locus controlling susceptibility to monophasic remitting/nonrelapsing experimental allergic encephalomyelitis. *J. Immunol.* **163:** 2262–2266.

Todd, J.A. 1999. From genome to aetiology in a multifactorial disease, type 1 diabetes. *BioEssays* **21:** 164–174.

Todd, J.A., Aitman, T.J., Cornall, R.J., Ghosh, S., Hall, J.R., Hearne, C.M., Knight, A.M., Love, J.M., McAleer, M.A., Prins, J.B., et al. 1991. Genetic analysis of autoimmune type 1 diabetes mellitus in mice. *Nature* **351:** 542–547.

Tukey, J.W. 1977. *Exploratory data analysis*. Wesley Publishing, Reading, MA.

Vyse, T.J. and Todd, J.A. 1996. Genetic analysis of autoimmune disease. *Cell* **85:** 311–318.

Ward, S.G. and Westwick, J. 1998. Chemokines: Understanding their role in T-lymphocyte biology. *Biochem. J.* **333:** 457–470.

Wicker, L.S., DeLarato, N.H., Pressey, A., and Peterson, L.B. 1993. Genetic control of diabetes and insulitis in the nonobese diabetic mouse: analysis of the NOD.H-2^b and B10.H-2^b strains. In *Molecular mechanisms of immunological self-recognition* (ed. Alt, F.W. and Vogel, H.J.) pp 173–181. Academic Press, New York.

Wicker, L.S., Todd, J.A., and Peterson, L.B. 1995. Genetic control of autoimmune diabetes in the NOD mouse. *Annu. Rev. Immunol.* **13:** 179–200.

Widmer, U., Yang, Z., van Deventer, S., Manogue, K.R., Sherry, B., and Cerami, A. 1991. Genomic structure of murine macrophage inflammatory protein-1 alpha and conservation of potential regulatory sequences with a human homolog, LD78. *J. Immunol.* **146:** 4031–4040.

Zhu, G., Spellman, P.T., Volpe, T., Brown, P.O., Botstein, D., Davis, T.N., and Futcher, B. 2000. Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature* **406:** 90–94.

Received September 6, 2001; accepted in revised form November 15, 2001.

Combining Mouse Congenic Strains and Microarray Gene Expression Analyses to Study a Complex Trait: The NOD Model of Type 1 Diabetes

Iain A. Eaves, Linda S. Wicker, Ghassan Ghandour, et al.

Genome Res. 2002 12: 232-243
Access the most recent version at doi:10.1101/gr.214102