CHAPTER

107

Genomic Approaches to the Host Response to Pathogens

M. Frances Shannon

INTRODUCTION

Following a period of relative disinterest in infectious disease research due to the enormous impact of vaccines and antibiotics on the spread of and mortality from these diseases, there is now renewed and growing interest in this area of research. This has been driven by several recent worldwide developments including: (i) the rising incidence of diseases such as Acquired Immune Deficiency Syndrome (AIDS) and antibiotic-resistant tuberculosis; (ii) antibiotic-resistant bacterial strains presenting a severe health threat in hospitals; (iii) the rapid spread of new pathogens such as Severe Acute Respiratory Syndrome (SARS); and (iv) the threat of bioterrorism. Indeed, nearly 25% of annual deaths worldwide are due to infectious disease (Morens et al., 2004). Thus, the need to develop new diagnostic methods, more effective vaccines and better therapeutic strategies is urgent.

In order to effectively deal with infectious disease threats, it is important to understand both the pathogen and the response of the host, since the outcome of infection is determined by complex host–pathogen interactions. Pathogens are initially detected by the surveillance cells of the innate immune system using cell surface receptors known as Toll-like receptors (TLRs) (reviewed in [Cook et al., 2004]). These TLRs recognize specific components of the pathogen, for example, bacterial lipopolysaccharide (LPS) or double-stranded (ds) RNA from viruses. While many cell types express TLRs, cells of the innate immune system

such as dendritic cells (DCs) and macrophages play particularly important roles in detecting and responding to pathogens. The response of these cells to a pathogen is determined by the specific pathogen component that interacts with the TLR and the specific TLR family member that is activated. Widespread changes in gene expression are detected following TLR activation and the activated cells produce a plethora of cytokines and chemokines that then activate the adaptive arm of the immune system. The specific cytokines and chemokines produced by the TLR-activated cells tailor the response of the adaptive immune system to deal with the specific pathogen (Cook et al., 2004). Thus, the initial host response to a pathogen through the TLRs determines the outcome of the infection. Host response to infection can be a double-edged sword in that sometimes the response itself can create an adverse outcome for the host. In addition, the aberrant response of the host to self instead of foreign pathogens can create severe pathologies involving chronic inflammatory and autoimmune diseases.

The urgent need to better understand host–pathogen interactions has come at a time when genomics and related technologies are expanding rapidly. The availability of complete genomic sequences of an expanding number of pathogens, the human and mouse genome sequences and the advent of genome-wide genotyping and gene expression profiling has opened up new avenues of investigation in the field.

The genotype of the pathogen plays a major role in the response of the host to infection with more virulent pathogenic

strains often possessing the capability to interfere with the host immune response (Fitzgerald and Musser, 2001; Kato–Maeda et al., 2001; Schoolnik, 2002). In addition, different individuals in a population can have very different responses to a genetically identical pathogen. While there are many complex reasons for this, it is clear that part of the differential response is governed by underlying genetic differences between individuals (Clementi and Di Gianantonio, 2006; Zhang and Zhang, 2006). Studies in mouse models of infection have clearly demonstrated that these genetic differences are complex and may involve more than one genetic locus for a given susceptibility or resistance trait (e.g., [Delahaye et al., 2006]). While there are some classic examples of genetic mutations affecting the response of the host to a pathogen (e.g., malaria and sickle cell mutations) there is much to be learned before the genetics of host susceptibility is fully understood. The advent of genome-wide genotyping using single nucleotide polymorphisms (SNPs) or microsatellite markers, leading to major advances in molecular epidemiology, will revolutionize our ability to determine the complexities of the genetic component of pathogen–host interactions (Weiss and Terwilliger, 2000).

It is well known that the cells of the host immune system are activated upon detection of a pathogen by TLRs as described above. This activation process includes widespread changes in the gene expression profile of the cells with hundreds of genes being either switched on or off in response to signals generated from the pathogen-detecting TLRs. The response of individual genes has been studied in minute detail for a handful of genes and while this has produced an understanding of some aspects of host response to infection it by no means gives us the total picture. Understanding the molecular response of the host to infection has been greatly improved by using microarray-based technologies and these technologies are opening up new diagnostic possibilities as well as presenting new therapeutic options (Aderem and Smith, 2004; Bryant et al., 2004; Feezor et al., 2005; Hedeler et al., 2006; Korth and Katze, 2002; Ng et al., 2006; Ricciardi-Castagnoli, 2005; Ricciardi-Castagnoli and Granucci, 2002; Smith and Bolouri, 2005).

This chapter will focus on two aspects of the host response to pathogens where major advances are being made using genomics approaches and will describe the future impact of these approaches on the development of diagnostics and therapeutics for infectious disease. These are (i) defining the basis of genetic susceptibility to infection and (ii) the definition of the system-wide molecular response to a pathogen.

GENETIC SUSCEPTIBILITY TO PATHOGENS

It is now relatively easy to map genes associated with genetic diseases that show a Mendelian pattern of inheritance. However, these diseases account for only a very small proportion of the human disease burden and many of the more common and fatal diseases have a complex etiology with many genetic and environmental contributions. While it is clear that most complex disease has a genetic component, defining that genetic component has been difficult to date since many complex diseases such as coronary heart disease, diabetes and others are polygenic with different genetic loci contributing in major or minor ways to disease susceptibility. In addition, in different populations or under different environmental conditions, distinct but overlapping sets of genetic loci are likely to contribute. The sequencing of the human genome and the genome-wide genetic and functional mapping that has followed have raised hopes of mapping the genetic component of complex disease and there are many large efforts around the world with this aim.

Studies in both animal models and human populations have shown that infectious disease and the response of the host to a specific infection also has a complex genetic component (Clementi and Di Gianantonio, 2006; Lipoldova and Demant, 2006; Marquet et al., 1996; Mira et al., 2004). Thus, inbred mouse models have been developed that clearly show a genetic component to susceptibility for specific pathogens and in some cases at least part of the underlying genetic reason has been defined (Beck et al., 2000; Mak et al., 2001; Rogner and Avner, 2003). Mapping the genetic components of susceptibility to infection in human populations has been much more difficult due to the large natural variation in humans, the polygenic nature of this trait and the low penetrance of many of the susceptibility alleles. For infectious disease, this is complicated even more by the complex nature of the environmental influences particularly the fact that these diseases, unlike other complex diseases, are transmissible in populations. However, a combination of animal and human population studies, combined with the latest genomic technologies, is beginning to unravel the issues of genetic susceptibility to infection.

The use of inbred and congenic strains of mice are well established systems for identifying susceptibility loci (Beck et al., 2000; Rogner and Avner, 2003). In recent years genetic manipulation of specific loci by deletion or mutation has provided many mouse models for screening (Mak et al., 2001). The use of ethylnitrosourea (ENU) mutagenesis to randomly create point mutations in the mouse genome has opened up a new forward genetics approach to identifying susceptibility loci (Papathanasiou and Goodnow, 2005). This chemical mutagen, when used at appropriate doses and at the correct stage of development, can introduce single point mutations into the mouse genome. By screening libraries of mutant mice for susceptibility to specific pathogens, it should be possible to identify genetic loci that dictate susceptibility or resistance to a range of pathogens on a large scale. It is relatively straightforward to identify a chromosomal region involved in susceptibility in these mouse strains by genotyping with microsatellite markers, but identifying the specific gene that is mutated is still very time-consuming. The speed with which this can be achieved depends on the presence of candidate genes within the chromosomal interval or the ability to resequence large amounts of DNA. The latter is becoming achievable with the advent of new rapid sequencing technologies and is set to revolutionize forward genetic

approaches to disease understanding (Bennett et al., 2005; Serre and Hudson, 2006).

The recent explosion in genetic information for the human genome including the complete genome sequence and detailed genetic and physical maps has increased our ability to find variation in the human genome and correlate it with disease. Family studies, especially twin studies, and population studies have clearly shown a genetic component to susceptibility to infectious disease (Frodsham and Hill, 2004; Lipoldova and Demant, 2006; Strunk and Burgner, 2006). Susceptibility generally follows a complex pattern of inheritance and there are two main methods of mapping and identifying genetic loci involved in such complex heredity. These are either association studies or linkage studies. Association studies involve screening populations for specific mutations in a candidate gene(s) in case–control studies or in family studies. This type of approach identified the link between Human Immunodeficiency Virus (HIV) resistance and the chemokine receptor, CCR5 described below (Dean et al., 1996; Samson et al., 1996). Genome–wide association studies although still quite expensive are now becoming feasible and being used to define linkage between specific markers and susceptibility. These linkage studies depend on the availability of markers and the density of these markers is rapidly increasing

with the large scale identification of new SNPs across the human genome. The selection of which SNPs to use and the large numbers of samples needed to generate statistically significant associations for low penetrance alleles are still challenges. The Haplotype MAP (HapMap) project is starting to identify haplotypes within different population groups and together with improvements in large scale genotyping technology and bioinformatics should be useful in studies of complex disease inheritance. Table 107.1 summarizes the best studied genetic susceptibility loci for response to different infectious agents in both mouse models and human studies.

One of the classical examples of genetic susceptibility to infection is the role of the hemaglobinopathies in the outcome of malaria infection (Patrinos et al., 2005). There are also certain chromosomal regions and families of genes that have attracted attention in terms of searching for susceptibility alleles or polymorphisms. Because the TLR family of receptors plays a major role in recognizing pathogens, it was speculated that genetic variation in these receptors or their signaling pathways might be responsible for some susceptibility phenotypes (reviewed in [Schroder and Schumann, 2005]). One of the best examples to date is the occurrence of a single polymorphism in the region of the human TLR4 gene encoding the extracellular domain of the

TABLE 107.1	List of well-described susceptibility loci for resistance or susceptibility to infectious disease.	
Pathogen	**Genes**	**References**
HIV	CCR5	Dean et al. (1996); Samson et al. (1996)
	HLA Class I	Hendel et al. (1999); Li et al. (2007); Selvaraj et al. (2006)
	CCR2	Magierowska et al. (1999); Su et al. (1999)
Malaria	Globin locus	Reviewed in Patrinos et al. (2005)
	HLA Class I	Migot-Nabias et al. (2001); Young et al. (2005); Reviewed in Hill (1996, 1999)
	TNF-α promoter	McGuire et al., (1994); Ubalee et al. (2001)
Leprosy	TLR2	Kang and Lee (2002); Bochud et al. (2003)
	HLA Class II	Shaw et al. (2001); Mehra et al. (1995)
	TNF-α promoter	Roy et al. (1997); Shaw et al. (2001)
Legionella	TLR5	Hawn et al. (2003); Merx et al. (2006)
Tuberculosis	HLA Class I	Lombard et al. (2006); Vijaya Lakshmi et al. (2006)
	NRAMP1/Slc11a1	Kusuhara et al. (2007); Li et al. (2006)
Typhoid fever	HLA Class II	Dunstan et al. (2001); Dharmana et al. (2002)
Leishmania	NRAMP1/Slc11a1	Bucheton et al. (2003); Mohamed et al. (2004)
	HLA	Reviewed in Lipoldova and Demant (2006)
	TNF	Bucheton et al. (2003)
	IFNgR1	Mohamed et al. (2003)
	IL-4	Mohamed et al. (2003)
Salmonella	NRAMP1/Slc11a1	Sebastiani et al. (1998)
Inhaled E. coli LPS	TLR4	Arbour et al. (2000); Feterowski et al. (2003)
Pyogenic bacteria	IRAK4	Picard et al. (2003)

This lists includes genes identified in both mouse and human studies.

receptor which confers reduced sensitivity to inhaled *Escherichia coli* (*E. coli*) LPS (Arbour et al., 2000). Interestingly, when septic shock patients were compared with a control group, these lower-responding alleles were found only in the septic shock group and these individuals had a higher incidence of Gram-negative bacterial infection (Feterowski et al., 2003). Such studies need further confirmation since there are also a number of studies that failed to find any linkage between TLR4 mutations and response to various infections (Schroder and Schumann, 2005). There is also enormous variation in the response of individuals to LPS even in the absence of TLR4 mutations implying that variation may occur in other components of the TLR4 signaling system. An example of this is the link between IRAK4 mutations and increased susceptibility to infection with pyogenic bacteria (Picard et al., 2003). Variation in other TLR genes has also been associated with disease susceptibility. For example, a mutation in the extracellular domain of TLR2 is linked to susceptibility to leprosy (Alcais et al., 2005) and a mutation in TLR5 increases susceptibility to *Legionella* (Hawn et al., 2003). Taken together these data support the idea that variation in the innate immune recognition of pathogens play an important part in governing susceptibility to an array of infectious diseases. However, caution needs to be exercised until larger population groups have been studied.

The extensive polymorphism at the chromosomal regions encoding major histocompatibility complex (MHC) proteins is thought to have arisen through natural selection in response to selective pressure from infectious disease. Although human leukocyte antigen (HLA) association with resistance or susceptibility to infectious disease has been difficult to identify because of the complex array of antigenic epitopes involved, a number of studies have implicated this locus in genetic susceptibility to infectious disease (Ghodke et al., 2005; Little and Parham, 1999). MHC molecules fall into two classes, Class I that present foreign antigens to CD8+ cytotoxic T cells and Class II that play a similar role for CD4+ helper T cells. Variation in specific Class I genes has been shown to confer susceptibility to pulmonary tuberculosis and to HIV whereas mutations in other Class I genes confer resistance to HIV and to severe malaria. Class II mutations that confer resistance to hepatitis B or hepatitis C have been identified and susceptibility to typhoid fever and leprosy are also associated with specific Class II mutations. Further molecular analysis of these and other associations may in the future have an impact on the development of new vaccines and immunotherapeutics.

To date the most successful manner of identifying susceptibility genes in human populations has been the candidate gene approach. Candidate genes have emerged from many sources including mouse genetic studies as well as biochemical and function dissection of the immune system. Once a candidate gene is identified, the chromosomal region spanning this gene in the human genome is then scanned for the occurrence of specific mutations or for functional polymorphisms in case–control studies across populations or in linkage studies in family groups. Such studies have identified a number of well described

susceptibility loci for infection with various pathogens. One of the most heralded example was the identification of a deletion in the chemokine receptor, CCR5, which was shown to confer resistance to HIV infection (Dean et al., 1996; Samson et al., 1996). Biochemically, this can be explained by the fact that CCR5 is a co-receptor for HIV on the surface of T cells (Dragic et al., 1996). A mutation in another chemokine receptor, CCR2, has also been shown to confer HIV resistance in certain Caucasian populations (O'Brien and Moore, 2000).

Some genes have been associated with susceptibility or resistance to multiple pathogens. For example, variation in the *NRAMP1/Slc11a1* gene is associated with susceptibility to *Leishmania* and to specific intracellular bacteria such as tuberculosis (Barton et al., 1999; Govoni et al., 1996; Lipoldova and Demant, 2006; Sebastiani et al., 1998). Mutations in the tumor necrosis factor (TNF) locus, mainly gene promoter mutations, have been linked with malaria and leprosy susceptibility (Lipoldova and Demant, 2006). Gene promoter or control region mutations have an impact on the level of protein produced from the gene rather than the function of the protein. This is an area of great interest but more difficult to study for several reasons, including the inability to identify control regions simply from sequence information and the complexity and flexibility of transcriptional control. A recent review detailing the genes associated with *Leishmania* susceptibility describes a number of genes that can affect disease outcome including the *interferon-gamma receptor type 1* (*IFNGR1*), the *interleukin-4 (IL-4)* gene and the *NRAMP-1/Slc11a1* gene (Lipoldova and Demant, 2006). These genes and others such as *interleukin-12* (*IL-12*) and its receptor are also linked with *Salmonella* and certain mycobacterial infections (Lipoldova and Demant, 2006). Thus, it is likely that variation in many genes can contribute to disturbing the finely balanced tuning of the immune system and lead to an altered response to a pathogen. It is clear from such studies that the same genes may be involved in susceptibility to an array of pathogens indicating a core immune response critical for any pathogen.

The identification of susceptibility loci for infection with various pathogens will aid in developing new diagnostic screens based on the detection of genetic variants in these loci. It could be envisaged that a person's susceptibility or resistance to a pathogen could be defined by a simple genotyping screen either prior to exposure to any pathogen or upon presentation with an infection. It may also be possible to determine the likely outcome of the infection through a genotyping screen. The definition of susceptibility loci will also contribute to our ability to develop new vaccines and therapeutics.

EXPLORING THE HOST RESPONSE THROUGH EXPRESSION PROFILING

The use of microarray technology to generate expression profiling data is becoming common place in biomedical research (reviewed in [Quackenbush, 2002; Sherlock, 2000]). This

technology allows the documentation of mRNA levels for thousands of genes from total RNA prepared from cells or tissue samples. The data obtained can be compared from sample to sample allowing the changes between samples to be documented and quantified. The "expression profile" for any cell or tissue is simply the list of genes whose expression can be detected using microarrays. The differences in the expression profile from one cell or tissue to the next or in cells treated in a specific manner is a surrogate measure of the cell/tissue phenotype and shows how that phenotype responds to its environment. Expression profiling is most useful when large datasets become available and when the data is combined with other data types and detailed bioinformatics studies. For example, using functional clustering of expression profiling data can help identify pathways that are important for a particular process and co-expression clustering combined with other technologies can help define regulatory networks within the cell.

Over the last 5–6 years, this technology has been applied to identifying the changes in gene expression that occur in response to infection by various pathogens (Aderem and Smith, 2004; Boyce et al., 2004; Bryant et al., 2004; Feezor et al., 2005; Foti et al., 2006; Jenner and Young, 2005; Korth and Katze, 2002; Korth et al., 2005; Ricciardi-Castagnoli and Granucci, 2002). To date there are more than 150 papers in the literature that describe gene expression changes that occur in response to infection with a plethora of pathogens and in many cell types (reviewed in [Jenner and Young, 2005]). Many of these are *in vitro* studies, taking specific cell types and infecting them with specific agents including bacteria, viruses, parasites and yeasts. In addition, cellular responses to bacterial components have also been documented, helping to identify pathogen-specific responses as well as determining the pathogenic component responsible for the major gene expression effects. Virulent or non-virulent strains of specific pathogens as well as mutant organisms have been used to determine the gene expression profile associated with a negative or positive clinical outcome. Few *in vivo* studies have been carried out and have shed light on the more complex responses seen in whole animals and helped to validate the *in vitro* data.

Identifying a Common Host Response to Infection

Several pioneering studies demonstrated that microarrays could be used to determine changes in the gene expression profile of cells in response to virus or bacterial infection (Boldrick et al., 2002; Gao et al., 2002; Huang et al., 2001; Nau et al., 2002). These studies paved the way for the analysis of the host response to a wide variety of infectious agents. The most significant of these studies compared the response of macrophages or DCs to a variety of infectious agents in a single study. In these studies a strong shared response to all infections, be they bacterial, viral or parasitic in nature, was identified. Not only was there commonality from one infectious agent to another but there was also some commonality across cell types. This expression signature has been interpreted as a general "alarm signal" for infection (reviewed in [Jenner and Young, 2005]). Studies of infection

with Gram-positive and Gram-negative bacteria also revealed a common expression signature in peripheral blood mononuclear cells. Recently, the Young lab has interrogated all of the publicly available expression profiling data related to the host response to infection (Jenner and Young, 2005). The dataset includes 785 experiments in cells ranging from macrophages and DCs to cells of the adaptive immune response, endothelial and epithelial cells and spanning a wide range of infecting agents. This meta-analysis revealed that a "common host response" can be detected across all of these cell types and infectious agents and show that although cells such as macrophages and DCs specialize in detecting infection, other cells of the body can mount the same "alarm response" as described above (Figure 107.1).

Not surprisingly this expression signature contains many genes associated with the immune system particularly those encoding inflammatory cytokines, chemokines and their receptors. However, some more surprising patterns of expression were also detected. It has been long known that interferon-stimulated genes (ISGs) are regulated by virus infection, but it has only recently been recognized that bacteria and other infecting agents can also illicit the interferon response. This was borne out in these meta-analyses where upregulation of an ISG set is observed across a broad range of infecting agents and cell types.

Not only do these cells change the expression of secreted factors and their receptors during infection but the intracellular milieu is also modified. Once again, there is a common pattern of change observed in all of these studies. The upregulation of signaling and transcription pathways that both augment and attenuate the immune response are observed leading to the interpretation that both positive and negative feedback loops operate within the cell to heighten or dampen the immune
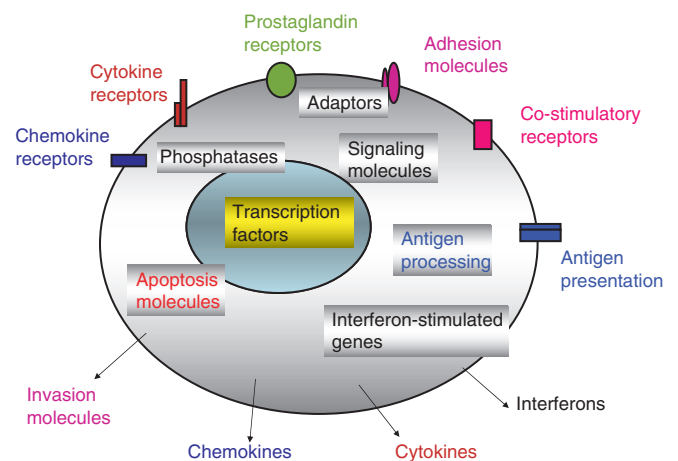

Figure 107.1 Summary of the common host response to infection. This figure is adapted from Jenner and Young (2005) and summarizes their meta-analysis of gene expression changes in response to infection with various pathogens or activation of a variety of cells with agents that mimic infection. Genes are grouped into general functional categories.

response. Temporal profiling can reveal extra layers of complexity and in one study a pro-inflammatory profile followed by an anti-inflammatory profile was identified in macrophages activated with LPS (Wells et al., 2005). Meta-analysis of temporal studies of activation through different TLRs also revealed that the inflammatory chemokines/cytokine signature was an early response while the ISG response was later, presumably reflecting the need for an indirect activation of the ISGs through interferon production (Jenner and Young, 2005). Changes in the expression level of genes involved in both activation and repression of apoptosis also fall under the "common signature" banner and this is interpreted as sending the cells into a state of high alert where apoptosis can be either initiated to eliminate infected cells or terminated if the infection resolves (Jenner and Young, 2005).

Although these *in vitro* studies have provided an overview of the response of isolated cell types to pathogenic infection, *in vivo* studies are needed to validate any of these results before application to clinical medicine. A number of animal models have been used to profile the host response to infection with a variety of agents. These studies are complicated especially if whole tissue samples are used in that changes in gene expression can result not only from genuine changes within the cells of the tissue but also from the recruitment especially of immune cells into the infected or inflamed tissue. Nevertheless, *in vivo* studies have, in some cases, shown good correlation with the expression profiles found from *in vitro* studies. For example, profiling the brains of mice infected with virulent Sindbis virus revealed that ISGs as well as inflammatory chemokines were upregulated (Johnston et al., 2001) and in terms of diagnosis or treatment the exact reason behind these changes in expression signature may be irrelevant. *In vivo* studies also have some advantages in that the gene expression profiles detected in infected tissues will often make more sense when combined with other physiological or cell biology data from studies of the infected host. Thus an "infection signature" would not only describe the altered gene expression of the immune cells that are recruited to the sites of infection but also would include changes in the gene expression of the resident cells of the tissue and may provide a more robust profile of the infection process for use in diagnostic applications.

What can be applied to clinical medicine from these studies? The ability to detect an "infection signature" using focused microarrays could potentially be used as a diagnostic tool. There would be an immediate need to identify a core set of genes with sufficiently robust changes in gene expression to form the basis of a diagnostic array. Arraying technology would need to be priced for diagnostic use and the technology would have to be deemed sufficiently robust to pass all the quality control requirements of a diagnostic laboratory. No doubt progress will be made toward these goals in the near future.

Pathogen-Specific Responses

In addition to the "common host response" described above, microarray studies have revealed that pathogen-specific responses also exist. This is not surprising since it has long been known that different pathogens induce distinct arms of the adaptive immune response. For example, distinct types of helper T cells are activated in response to bacterial and viral infection compared to parasite infection and DC products such as cytokines and chemokines control the differentiation of T helper cell into the correct response type (Mosmann et al., 2005). Thus, the fact that cells of the innate immune system display a pathogen-specific transcriptional response as well as a general alarm signal helps to dictate the subsequent immune response.

Different TLRs are involved in recognizing and responding to different pathogens. For example, TLR2 is responsible for activation in response to Gram-positive bacteria while TLR4 responds to LPS, a component of Gram-negative bacteria (Beutler and Rietschel, 2003; Cook et al., 2004). TLR3 responds to dsRNA and thus dictates the viral immune response for many dsRNA viruses (Beutler and Rietschel, 2003; Cook et al., 2004). Studies using bacterial components such as LPS and flagellin as well as dsRNA have revealed that each TLR induces a specific as well as general transcriptional response. The transcriptional response of macrophages and peripheral blood mononuclear cells is more robust in response to Gram-negative (TLR4) compared with Gram-positive (TLR2) bacteria and the ISG response is considerably reduced for the Gram-positive expression signature (Boldrick et al., 2002; Nau et al., 2002). These differences are further observed when bacterial components are used to activate cells. For example, LPS from Gram-negative bacteria, a specific ligand for TLR4, can activate the ISG profile but TLR2 ligands such as LTA and MDP cannot (Jenner and Young, 2005). Not only the activation of specific gene sets but also the strength of the specific response signature may be important for the immune detection of the type of pathogen involved. *E. coli* infection of DCs strongly upregulates the chemokines/cytokine inflammatory cluster whereas infection with influenza or other single stranded (ss) RNA viruses (through TLR7) has a weaker ability to regulate this cluster but a stronger ability to regulate the ISG signature (Huang et al., 2001; Lund et al., 2004).

These types of results raise the possibility that the diagnosis of the type of pathogen involved in an infection would be helped by the development of customized microarrays that could distinguish the gene expression profiles elicited by particular pathogens. Additionally, arrays that also detect RNAs produced by the pathogen may be even more significant as a diagnostic tool.

Determining the Outcome of the Infection

An infecting agent can either be cleared from the body by the immune system mounting an appropriate response or cause severe or terminal pathology. The outcome depends on a multiplicity of events ranging from the genotype of the host, that is, whether the host displays a resistant or susceptible phenotype, the genotype of the infectious agent, that is, virulent or non-virulent strains and many other less defined environmental factors. Can genomic approaches be used to determine the outcome of infection? Clearly, as discussed above, the technology is developing to define susceptible and resistant host genotypes especially in animal models of infection but the ability to do this routinely in human populations is some way into the future. Given the smaller

genomes of pathogenic organisms, defining virulence genotypes is progressing at a faster rate (Chan, 2003; Dorrell et al., 2005; Fitzgerald and Musser, 2001; Kato-Maeda et al., 2001; MacFarlane et al., 2005; Schoolnik, 2002; Zhang and Zhang, 2006).

Expression profiling studies have been used to investigate the differences in the host response to pathogenic and non-pathogenic strains of specific infectious agents. In one example, mice infected with a pathogenic strain of pneumonia virus upregulated the expected inflammatory chemokines/cytokine profile as well as the ISG profile but an attenuated strain of the same virus could not, although the virus replicated in the lungs of these mice to the same degree (Domachowske et al., 2001). Temporal profiling of the infection process in animals will also help to define the expression signatures associated with the ability of the host to clear specific pathogens.

GENETICAL GENOMICS AND SYSTEMS BIOLOGY: THE NEW FRONTIERS

While considerable information and progress is being made with the approaches described above, there is now much interest in combining both a genetic screening approach with an expression profiling approach in a new area of investigation being dubbed "genetical genomics" (reviewed in [Cook et al., 2004; de Koning et al., 2005; Schadt et al., 2003]). The rationale behind this combination is twofold: first, expression profiling data can be used to identify those genes in a chromosomal region previously identified from genetic approaches, whose expression is altered between two different genotypes. These genes then immediately become strong candidates for a susceptibility gene at that genetic locus. In this way, mutations that affect the expression but not the function of a protein can also be identified. Second, expression profiling combined with genotyping can identify functional pathways that can be affected by a single mutation. For example, if a mutation affects a transcription factor that controls expression of a group of functionally related genes then not only the transcription factor but the entire pathway will be identified opening up better opportunities to design new therapeutics.

In a prelude to these types of approaches, Hume and colleagues (Wells et al., 2003) carried out expression profiling studies on LPS-stimulated macrophages from various mouse strains that show differential response to infection. These studies showed that while there was a common core response to TLR4 activation, each strain of mice showed a unique gene expression program implicating many different genetic loci in this variable response. Mapping the genetic loci responsible for these different transcriptional responses will help to shed light on genetic loci that may play a role in human susceptibility to infection.

These type of approaches move into the realms of "systems biology" which is the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions, focusing on all the components and the interactions among them, as part of one system. Given the complexity of the immune response to infection, ultimately this type of approach may provide more answers.

APPLICATION TO CLINICAL PRACTICE

Although genomic approaches to understanding the host response to infectious disease are still very much at the developmental stage, it is likely that such approaches will be applied to diagnosis in the near future. In fact there are already some examples of the application of these approaches. A number of studies have been described using oligonucleotide-based arrays for the detection of multiple pathogens in a single experiment (Campbell and Ghazal, 2004; Dietel and Sers, 2006; Hervas, 2004; Kostrzynska and Bachand, 2006; Pompe et al., 2005). Recent studies have investigated the use of expression profiling of whole blood to determine human subjects at risk of recurrent tuberculosis (Mistry et al., 2007) and to identify high and low responders to lipopolysaccharide (Wurfel et al., 2005). A study in macaques has used whole blood expression profiling to examine molecular signatures of influenza infection (Baas et al., 2006). There are many efforts currently to optimize expression profiling for biomarker discovery and for clinical trials (Debey et al., 2006; Shou et al., 2005; Zheng et al., 2006). There are still many hurdles to be over come in terms of application to the clinical setting including standardization of sample preparation, robustness of the microarray platform and data analysis, cost quality assurance and so on (Abdullah-Sayani et al., 2006). The introduction of genetic or genomic screening into the clinical setting for infectious disease will no doubt occur in the near future.

ACKNOWLEDGEMENT

REFERENCES

Abdullah-Sayani, A., Bueno-de-Mesquita, J.M. et al. (2006). Technology insight: Tuning into the genetic orchestra using microarrays–limitations of DNA microarrays in clinical practice. *Nat Clin Pract Oncol* 3(9), 501–516.

Aderem, A. and Smith, K.D. (2004). A systems approach to dissecting immunity and inflammation. *Semin Immunol* 16(1), 55–67.

Alcais, A., Mira, M. et al. (2005). Genetic dissection of immunity in leprosy. *Curr Opin Immunol* 17(1), 44–48.

Arbour, N.C., Lorenz, E. et al. (2000). TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet* 25(2), 187–191.

Baas, T., Baskin, C.R. et al. (2006). Integrated molecular signature of disease: Analysis of influenza virus-infected macaques

through functional genomics and proteomics. *J Virol* 80(21), 10813–10828.

Barton, C.H., Biggs, T.E. et al. (1999). Nramp1: A link between intracellular iron transport and innate resistance to intracellular pathogens. *J Leukoc Biol* 66(5), 757–762.

Beck, J.A., Lloyd, S. et al. (2000). Genealogies of mouse inbred strains. *Nat Genet* 24(1), 23–25.

Bennett, S.T., Barnes, C. et al. (2005). Toward the 1,000 dollars human genome. *Pharmacogenomics* 6(4), 373–382.

Beutler, B. and Rietschel, E.T. (2003). Innate immune sensing and its roots: The story of endotoxin. *Nat Rev Immunol* 3(2), 169–176.

Bochud, P.Y., Hawn, T.R. et al. (2003). Cutting edge: A Toll-like receptor 2 polymorphism that is associated with lepromatous leprosy is unable to mediate mycobacterial signaling. *J Immunol* 170(7), 3451–3454.

Boldrick, J.C., Alizadeh, A.A. et al. (2002). Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proc Natl Acad Sci USA* 99(2), 972–977.

Boyce, J.D., Cullen, P.A. et al. (2004). Genomic-scale analysis of bacterial gene and protein expression in the host. *Emerg Infect Dis* 10(8), 1357–1362.

Bryant, P.A., Venter, D. et al. (2004). Chips with everything: DNA microarrays in infectious diseases. *Lancet Infect Dis* 4(2), 100–111.

Bucheton, B., Abel, L. et al. (2003). Genetic control of visceral l eishmaniasis in a Sudanese population: Candidate gene testing indicates a linkage to the NRAMP1 region. *Genes Immun* 4(2), 104–109.

Campbell, C.J. and Ghazal, P. (2004). Molecular signatures for diagnosis of infection: Application of microarray technology. *J Appl Microbiol* 96(1), 18–23.

Chan, V.L. (2003). Bacterial genomes and infectious diseases. *Pediatr Res* 54(1), 1–7.

Clementi, M. and Di Gianantonio, E. (2006). Genetic susceptibility to infectious diseases. *Reprod Toxicol* 21(4), 345–349.

Cook, D.N., Pisetsky, D.S. et al. (2004). Toll-like receptors in the pathogenesis of human disease. *Nat Immunol* 5(10), 975–979.

Cook, D.N., Wang, S. et al. (2004). Genetic regulation of endotoxin-induced airway disease. *Genomics* 83(6), 961–969.

de Koning, D.J., Carlborg, O. et al. (2005). The genetic dissection of immune response using gene-expression studies and genome mapping. *Vet Immunol Immunopathol* 105(3–4), 343–352.

Dean, M., Carrington, M. et al. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* 273(5283), 1856–1862.

Debey, S., Zander, T. et al. (2006). A highly standardized, robust, and cost-effective method for genome-wide transcriptome analysis of peripheral blood applicable to large-scale clinical trials. *Genomics* 87(5), 653–664.

Delahaye, N.F., Coltel, N. et al. (2006). Gene-expression profiling discriminates between cerebral malaria (CM)-susceptible mice and CM-resistant mice. *J Infect Dis* 193(2), 312–321.

Dharmana, E., Joosten, I. et al. (2002). HLA-DRB1*12 is associated with protection against complicated typhoid fever, independent of tumour necrosis factor alpha. *Eur J Immunogenet* 29(4), 297–300.

Dietel, M. and Sers, C. (2006). Personalized medicine and development of targeted therapies: The upcoming challenge for d iagnostic molecular pathology. A review. *Virchows Arch* 448(6), 744–755.

Domachowske, J.B., Bonville, C.A. et al. (2001). Glucocorticoid administration accelerates mortality of pneumovirus-infected mice. *J Infect Dis* 184(12), 1518–1523.

Dorrell, N., Hinchliffe, S.J. et al. (2005). Comparative phylogenomics of pathogenic bacteria by microarray analysis. *Curr Opin Microbiol* 8(5), 620–626.

Dragic, T., Litwin, V. et al. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 381(6584), 667–673.

Dunstan, S.J., Stephens, H.A. et al. (2001). Genes of the class II and class III major histocompatibility complex are associated with typhoid fever in Vietnam. *J Infect Dis* 183(2), 261–268.

Feezor, R.J., Cheng, A. et al. (2005). Functional genomics and gene expression profiling in sepsis: Beyond class prediction. *Clin Infect Dis* 41(Suppl 7), S427–sS435.

Feterowski, C., Emmanuilidis, K. et al. (2003). Effects of functional Toll-like receptor-4 mutations on the immune response to human and experimental sepsis. *Immunology* 109(3), 426–431.

Fitzgerald, J.R. and Musser, J.M. (2001). Evolutionary genomics of pathogenic bacteria. *Trends Microbiol* 9(11), 547–553.

Foti, M., Granucci, F. et al. (2006). Dendritic cells in pathogen recognition and induction of immune responses: A functional genomics approach. *J Leukoc Biol* 79(5), 913–916.

Frodsham, A.J. and Hill, A.V. (2004). Genetics of infectious diseases. *Hum Mol Genet* 13(Spec No 2), R187–R194.

Gao, J.J., Diesl, V. et al. (2002). Regulation of gene expression in mouse macrophages stimulated with bacterial CpG-DNA and lipopolysaccharide. *J Leukoc Biol* 72(6), 1234–1245.

Ghodke, Y., Joshi, K. et al. (2005). HLA and disease. *Eur J Epidemiol* 20(6), 475–488.

Govoni, G., Vidal, S. et al. (1996). The Bcg/Ity/Lsh locus: genetic transfer of resistance to infections in C57BL/6J mice transgenic for the Nramp1 Gly169 allele. *Infect Immun* 64(8), 2923–2929.

Hawn, T.R., Verbon, A. et al. (2003). A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. *J Exp Med* 198(10), 1563–1572.

Hedeler, C., Paton, N.W. et al. (2006). A classification of tasks for the systematic study of immune response using functional genomics data. *Parasitology* 132(Pt 2), 157–167.

Hendel, H., Caillat-Zucman, S. et al. (1999). New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J Immunol* 162(11), 6942–6946.

Hervas, F. (2004). Chip-mediated techniques: How close are we to generalised use in the infectious disease clinic?. *Clin Microbiol Infect* 10(10), 865–867.

Hill, A.V. (1996). Genetic susceptibility to malaria and other infectious diseases: From the MHC to the whole genome. *Parasitology* 112(Suppl), S75–S84.

Hill, A.V. (1999). The immunogenetics of resistance to malaria. *Proc Assoc Am Physicians* 111(4), 272–277.

Huang, Q., Liu, D. et al. (2001). The plasticity of dendritic cell responses to pathogens and their components. *Science* 294(5543), 870–875.

Jenner, R.G. and Young, R.A. (2005). Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* 3(4), 281–294.

Johnston, C., Jiang, W. et al. (2001). Identification of genes involved in the host response to neurovirulent alphavirus infection. *J Virol* 75(21), 10431–10445.

Kang, T.J., Lee, S.B. et al. (2002). A polymorphism in the toll-like receptor 2 is associated with IL-12 production from monocyte in lepromatous leprosy. *Cytokine* 20(2), 56–62.

Kato-Maeda, M., Gao, Q. et al. (2001). Microarray analysis of pathogens and their interaction with hosts. *Cell Microbiol* 3(11), 713–719.

Korth, M.J., Kash, J.C. et al. (2005). Virus infection and the interferon response: A global view through functional genomics. *Methods Mol Med* 116, 37–55.

Korth, M.J. and Katze, M.G. (2002). Unlocking the mysteries of virus-host interactions: Does functional genomics hold the key?. *Ann NY Acad Sci* 975, 160–168.

Kostrzynska, M. and Bachand, A. (2006). Application of DNA microarray technology for detection, identification, and characterization of food-borne pathogens. *Can J Microbiol* 52(1), 1–8.

Kusuhara, K., Yamamoto, K. et al. (2007). Association of IL12RB1 polymorphisms with susceptibility to and severity of tuberculosis in Japanese: A gene-based association analysis of 21 candidate genes. *Int J Immunogenet* 34(1), 35–44.

Li, H.T., Zhang, T.T. et al. (2006). SLC11A1 (formerly NRAMP1) gene polymorphisms and tuberculosis susceptibility: A meta-analysis. *Int J Tuberc Lung Dis* 10(1), 3–12.

Li, S., Jiao, H. et al. (2007). Human leukocyte antigen class I and class II allele frequencies and HIV-1 infection associations in a Chinese cohort. *J Acquir Immune Defic Syndr* 44(2), 121–131.

Lipoldova, M. and Demant, P. (2006). Genetic susceptibility to infectious disease: Lessons from mouse models of leishmaniasis. *Nat Rev Genet* 7(4), 294–305.

Little, A.M. and Parham, P. (1999). Polymorphism and evolution of HLA class I and II genes and molecules. *Rev Immunogenet* 1(1), 105–123.

Lombard, Z., Dalton, D.L. et al. (2006). Association of HLA-DR, -DQ, and vitamin D receptor alleles and haplotypes with tuberculosis in the Venda of South Africa. *Hum Immunol* 67(8), 643–654.

Lund, J.M., Alexopoulou, L. et al. (2004). Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci USA* 101(15), 5598–5603.

MacFarlane, R., Bhattacharya, D. et al. (2005). Genomic DNA microarrays for Entamoeba histolytica: Applications for use in expression profiling and strain genotyping. *Exp Parasitol* 110(3), 196–202.

Magierowska, M., Theodorou, I. et al. (1999). Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. *Blood* 93(3), 936–941.

Mak, T.W., Penninger, J.M. et al. (2001). Knockout mice: A paradigm shift in modern immunology. *Nat Rev Immunol* 1(1), 11–19.

Marquet, S., Abel, L. et al. (1996). Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31-q33. *Nat Genet* 14(2), 181–184.

McGuire, W., Hill, A.V. et al. (1994). Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature* 371(6497), 508–510.

Mehra, N.K., Rajalingam, R. et al. (1995). Variants of HLA-DR2/DR51 group haplotypes and susceptibility to tuberculoid leprosy and pulmonary tuberculosis in Asian Indians. *Int J Lepr Other Mycobact Dis* 63(2), 241–248.

Merx, S., Zimmer, W. et al. (2006). Characterization and functional investigation of single nucleotide polymorphisms (SNPs) in the human TLR5 gene. *Hum Mutat* 27(3), 293.

Migot-Nabias, F., Luty, A.J. et al. (2001). HLA alleles in relation to specific immunity to liver stage antigen-1 from plasmodium falciparum in Gabon. *Genes Immun* 2(1), 4–10.

Mira, M.T., Alcais, A. et al. (2004). Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 427(6975), 636–640.

Mistry, R., Cliff, J.M. et al. (2007). Gene-expression patterns in whole blood identify subjects at risk for recurrent tuberculosis. *J Infect Dis* 195(3), 357–365.

Mohamed, H.S., Ibrahim, M.E. et al. (2003). Genetic susceptibility to visceral leishmaniasis in The Sudan: Linkage and association with IL4 and IFNGR1. *Genes Immun* 4(5), 351–355.

Mohamed, H.S., Ibrahim, M.E. et al. (2004). SLC11A1 (formerly NRAMP1) and susceptibility to visceral leishmaniasis in The Sudan. *Eur J Hum Genet* 12(1), 66–74.

Morens, D.M., Folkers, G.K. et al. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature* 430(6996), 242–249.

Mosmann, T.R., Cherwinski, H. et al. (2005). Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. 1986. *J Immunol* 175(1), 5–14.

Nau, G.J., Richmond, J.F. et al. (2002). Human macrophage activation programs induced by bacterial pathogens. *Proc Natl Acad Sci USA* 99(3), 1503–1508.

Ng, A., Bursteinas, B. et al. (2006). pSTIING: A "systems" approach towards integrating signalling pathways, interaction and transcriptional regulatory networks in inflammation and cancer. *Nucleic Acids Res* 34(Database issue), D527–D534.

O'Brien, S.J. and Moore, J.P. (2000). The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol Rev* 177, 99–111.

Papathanasiou, P. and Goodnow, C.C. (2005). Connecting mammalian genome with phenome by ENU mouse mutagenesis: Gene combinations specifying the immune system. *Annu Rev Genet* 39, 241–262.

Patrinos, G.P., Kollia, P. et al. (2005). Molecular diagnosis of inherited disorders: Lessons from hemoglobinopathies. *Hum Mutat* 26(5), 399–412.

Picard, C., Puel, A. et al. (2003). Pyogenic bacterial infections in humans with IRAK-4 deficiency. *Science* 299(5615), 2076–2079.

Pompe, S., Simon, J. et al. (2005). Future trends and challenges in pathogenomics. A Foresight study. *EMBO Rep* 6(7), 600–605.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat Genet* 32(Suppl), 496–501.

Ricciardi-Castagnoli, P. (2005). Genomic signatures of immune and inflammatory responses. *J Pediatr Gastroenterol Nutr* 40(Suppl 1), S45–S46.

Ricciardi-Castagnoli, P. and Granucci, F. (2002). Opinion: Interpretation of the complexity of innate immune responses by functional genomics. *Nat Rev Immunol* 2(11), 881–889.

Rogner, U.C. and Avner, P. (2003). Congenic mice: Cutting tools for complex immune disorders. *Nat Rev Immunol* 3(3), 243–252.

Roy, S., McGuire, W. et al. (1997). Tumor necrosis factor promoter polymorphism and susceptibility to lepromatous leprosy. *J Infect Dis* 176(2), 530–532.

Samson, M., Libert, F. et al. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382(6593), 722–725.

Schadt, E.E., Monks, S.A. et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422(6929), 297–302.

Schoolnik, G.K. (2002). Functional and comparative genomics of pathogenic bacteria. *Curr Opin Microbiol* 5(1), 20–26.

Schroder, N.W. and Schumann, R.R. (2005). Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infect Dis* 5(3), 156–164.

Sebastiani, G., Olien, L. et al. (1998). Mapping of genetic modulators of natural resistance to infection with *Salmonella typhimurium* in wild-derived mice. *Genomics* 47(2), 180–186.

Selvaraj, P., Swaminathan, S. et al. (2006). Association of human leukocyte antigen-A11 with resistance and B40 and DR2 with susceptibility to HIV-1 infection in south India. *J Acquir Immune Defic Syndr* 43(4), 497–499.

Serre, D. and Hudson, T.J. (2006). Resources for genetic variation studies. *Annu Rev Genomics Hum Genet.* 7, 443–457.

Shaw, M.A., Donaldson, I.J. et al. (2001). Association and linkage of leprosy phenotypes with HLA class II and tumour necrosis factor genes. *Genes Immun* 2(4), 196–204.

Sherlock, G. (2000). Analysis of large-scale gene expression data. *Curr Opin Immunol* 12(2), 201–205.

Shou, J., Dotson, C. et al. (2005). Optimized blood cell profiling method for genomic biomarker discovery using high-density microarray. *Biomarkers* 10(4), 310–320.

Smith, K.D. and Bolouri, H. (2005). Dissecting innate immune responses with the tools of systems biology. *Curr Opin Immunol* 17(1), 49–54.

Strunk, T. and Burgner, D. (2006). Genetic susceptibility to neonatal infection. *Curr Opin Infect Dis* 19(3), 259–263.

Ubalee, R., Suzuki, F. et al. (2001). Strong association of a tumor necrosis factor-alpha promoter allele with cerebral malaria in Myanmar. *Tissue Antigens* 58(6), 407–410.

Weiss, K.M. and Terwilliger, J.D. (2000). How many diseases does it take to map a gene with SNPs?. *Nat Genet* 26(2), 151–157.

Wells, C.A., Ravasi, T. et al. (2003). Genetic control of the innate immune response. *BMC Immunol* 4, 5.

Wells, C.A., Ravasi, T. et al. (2005). Inflammation suppressor genes: Please switch out all the lights. *J Leukoc Biol* 78(1), 9–13.

Wurfel, M.M., Park, W.Y. et al. (2005). Identification of high and low responders to lipopolysaccharide in normal subjects: An unbiased approach to identify modulators of innate immunity. *J Immunol* 175(4), 2570–2578.

Young, K., Frodsham, A. et al. (2005). Inverse associations of human leukocyte antigen and malaria parasite types in two West African populations. *Infect Immun* 73(2), 953–955.

Zhang, R. and Zhang, C.T. (2006). The impact of comparative genomics on infectious disease research. *Microbes Infect* 8(6), 1613–1622.

Zheng, Z., Luo, Y. et al. (2006). Sensitive and quantitative measurement of gene expression directly from a small amount of whole blood. *Clin Chem* 52(7), 1294–1302.

RECOMMENDED RESOURCES

Journals

Jenner, R.G., Young, R.A. (2005). Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* 3(4), 281–294. This review describes a meta-analysis of expression profiling data from the literature of cells infected with different pathogens or treated with pathogenic components.

Lipoldova, M., Demant, P. (2006). Genetic susceptibility to infectious disease: Lessons from mouse models of leishmaniasis. *Nat Rev Genet* 7(4), 294–305. This paper reviews the literature on the genetic susceptibility to infection with *Leishmania* and compares susceptibility loci to those identified for other infections.

Cook, D.N., Pisetsky, D.S., Schwartz, D.A. (2004). Toll-like receptors in the pathogenesis of human disease. *Nat Immunol* 5(10), 975–979. This review describes the role of Toll-like receptors in detection of pathogens and summarizes their involvement in infectious disease susceptibility.

Websites

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene

UniGene is an Organized View of the Transcriptome. Each UniGene entry is a set of transcript sequences that appear to come from the same transcription locus (gene or expressed pseudogene), together with information on protein similarities, gene expression, cDNA clone reagents and genomic location.

http://www.ncbi.nih.gov/entrez/query.fcgi?db=OMIM

Online Mendelian Inheritance in Man is a database catalog of human genes and genetic disorders.

http://www.genome.jp/kegg/ Kyoto Encyclopedia of Genes and Genomes.

http://pstiing.licr.org/

pSTIING (Protein, Signaling, Transcriptional Interactions and Inflammation Networks Gateway) is a publicly accessible knowledge-base about protein–protein, protein–lipid, protein–small molecules, ligand–receptor interactions, receptor–cell type information, transcriptional regulatory and signal transduction modules relevant to inflammation, cell migration and tumorigenesis.

http://www.genmapp.org/

Gene Map Annotator and Pathway Profiler is a computer application designed to visualize gene expression data on maps representing biological pathways and groupings of genes.

http://www.ensembl.org/index.html

Ensembl is a joint project between EMBL-EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.

http://www.informatics.jax.org/

MGD includes information on mouse genetic markers, molecular clones (probes, primers and YACs), phenotypes, sequences, comparative mapping data, graphical displays of linkage, cytogenetic and physical maps, experimental mapping data, as well as strain distribution patterns for recombinant inbred strains (RIs) and cross haplotypes.

http://www.geneontology.org/

The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism.

http://www.ncbi.nlm.nih.gov/geo/

Gene Expression Omnibus is a gene expression/molecular abundance repository supporting MIAME compliant microarray data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.

http://www.genome-www5.stanford.edu/

The Stanford MicroArray Database stores raw and normalized data from microarray experiments, and provides data retrieval, analysis and visualization.

http://www.expression.microslu.washington.edu/expression/index.html

Public Microarray Data Download Site powered by Expres-sion Array Manager.

http://www.ncbi.nlm.nih.gov/projects/SNP/

The National Center for Biotechnology Information has established the Single Nucleotide Polymorphism (dbSNP) database to serve as a central repository for both single base nucleotide substitutions and short deletion and insertion polymorphisms.