POLYMORPHISMS OF THE IMMUNE RESPONSE GENES: SELECTED BIOLOGICAL, METHODICAL AND MEDICAL ASPECTS

Zuzana Kubistova, Frantisek Mrazek, Martin Petrek*

Laboratory of Immunogenomics and Immunoproteomics, Department of Immunology, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic
e-mail: martin.petrek@fnol.cz

Received: October 10, 2008; Accepted: April 15, 2009

Key words: Immune response genes/Gene polymorphism/Susceptibility to diseases/Complex diseases

Background: Variants of the immune response genes (IRG) are considered a potential source of interindividual differences in both innate and adaptive immune responses. A large number of gene polymorphisms have been reported as alternative forms of the IRG nucleotide sequence with important functional consequences for the structure/expression of immune response molecules. Accordingly, IRG polymorphisms are considered responsible for various monogenic diseases. They may also affect individual predisposition to complex diseases or modify their clinical course.

Methods and Results: In this review we define IRG polymorphism including its potential functionality. Common approaches used for the investigation of IRG polymorphisms are next briefly described. We then review current approaches (including genome-wide studies) for assessing the importance of particular IRG variants in the susceptibility to and clinical course of complex diseases. Finally, based on our own experience and on the literature, we illustrate current knowledge of the genetic component of two selected complex diseases (sarcoidosis and coronary artery disease).

Conclusions: Despite major advances in genotyping technology and general knowledge of the implications of IRG in the susceptibility to complex diseases, the potential clinical application of these approaches still faces major challenges.

INTRODUCTION

The immune system plays a critical role in an individual’s autonomy and ability to recognise its own tissues from foreign, potentially dangerous entities (microbes, tumour cells etc.) (ref.1). The effector phase of the immune response is aimed at eliminating non-self entities from the body. It has been postulated that the immune response is not uniform or constantly effective throughout the population. Individual patterns of immune response are thought to be associated with variability (polymorphisms) of the immune response genes (IRG) (ref.2). As immune molecules are implicated in the pathogenesis of a large number of human diseases, variations in the DNA sequences of the IRG can affect whether/how humans develop these diseases and, respond to pathogens, chemicals, drugs, etc. In this review, we briefly define IRG polymorphisms, their functional relevance and techniques for their genotyping. Further, the approaches currently used for assessing the potential role of IRG polymorphisms in diseases are discussed. Finally, examples of earlier reported association of IRG polymorphisms to genetic susceptibility to complex diseases are included.

2. Characteristics of gene polymorphisms

2.1 Immune response genes and their variation

Immune response genes (IRG) may be defined as genes encoding for molecules and mediators implicated in both innate and adaptive immune response. The spectrum of IRG is wide and highly complex – from receptors of innate immunity (e.g. Toll-like receptors – TLRs), through cytokines and their receptors to histocompatibility antigens (MHC) and immunoglobulins3. Further, the genes for a large number of intracellular signalling molecules participating in the transfer of information may also be considered genes associated with the immune response.

Importantly, all general knowledge of the human genome can also apply to IRG. IRGs thus have been shown to be polymorphic; they exist normally in several variants (alleles) (ref.4). However, the level of polymorphism differs distinctly for particular groups or types of IRG. This is most probably related to their function. For example, although the coding sequence of pleiotropic cytokines (TNF, IL-1) is highly conservative, an extreme population variability has been observed in the genes encoding for molecules associated with antigen recognition (MHC, T-cell receptors – TCR, immunoglobulins).

2.2 Definition and structural types of gene (DNA) polymorphism

In general, gene (DNA) polymorphism is defined as two or more alternate forms (alleles) of a chromosomal locus that differ in nucleotide sequence or have variable numbers of repeated nucleotide units. Rare variations are not classified as polymorphisms – the criterion decided upon is that the major allele has a frequency of 99 % or less.
2.2.1 Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation. There are more than 10 million SNP polymorphisms in public databases. A SNP is a single base pair change (substitution/deletion/insertion of one nucleotide) at a specific locus, usually consisting of two alleles. Within a population, a minor allele frequency can be assigned for each SNP as the ratio of chromosomes in the population carrying the less common variant to those with the more common variant. In the past, single nucleotide polymorphisms with a minor allele frequency of $\geq 1\%$ were labelled “SNP”. It is important to note that there are marked differences between human populations in terms of the distribution of particular SNP variants. For this reason, an SNP allele that is common in one geographical or ethnic group may be much rarer in another.

2.2.2 Variable number of tandem repeats and copy number variation

The repeated DNA sequences, that occurrence in the human genome is highly frequent, come in various sizes and are classified according to the length of the repeat units, the number of contiguous repeat units, and/or the overall length of the repeat region. DNA regions with short repeat units (usually 2-6 bp in length) are called Short Tandem Repeats (STR), regions with longer repeat units are commonly called Variable Number of Tandem Repeats. Finally, variability in the number of whole functional gene units (exons, genes) is known as Copy Number Variation (CNV).

A short tandem repeat (STR) in DNA is a class of polymorphisms that occurs when a pattern of two or more nucleotides are repeated and the repeated sequences are directly adjacent to each other. STRs are most frequently found surrounding the chromosomal centromere (the structural center of the chromosomes). STRs have several proven benefits that make them especially suitable for human identification. By examining several STR loci and counting how many repeats of a specific STR sequence there are at a given locus, it is possible to create a unique genetic profile of an individual. There are currently over 10 000 published STR sequences in the human genome. STR analysis has become the prevalent method for determining genetic profiles in forensic cases. The “copy number variants” is a type of polymorphism characterised by differences in number of copies of a particular gene in the genotype of an individual. Recent evidence shows that the gene copy number can be elevated in cancer cells. Elevating the gene copy number of a particular gene can increase the expression of the protein that it encodes.

2.3 Functional DNA polymorphisms – effects on structure and expression of the immune response molecules

“Functional gene (DNA) polymorphism” is a term reserved for polymorphisms that change in any way, the expression and/or structure of the gene product (e.g. protein, specialized RNA). In general, polymorphisms located inside the coding sequence of the gene may encode for an alternative amino acid in the primary structure of the protein (a different polypeptide sequence is produced) or may cause failure in the non-functional protein (especially deletions/insertions of a nucleotide). These SNPs are called non-synonymous. SNPs within a coding sequence will not necessarily change the amino acid sequence of the protein that is produced, due to degeneration of the genetic code; an SNP in which both forms lead to the same polypeptide sequence is termed synonymous.

SNPs that are not in protein coding regions may still have consequences for transcript splicing / stability, transcription factor binding or the sequence of non-coding RNA. Variants in these “non-coding” regions may, in general, affect significantly baseline and/or induced expression of the gene and produce alternatively spliced variants. Numerous polymorphisms of non coding gene regions have already been shown to be associated with the deregulation of the encoded immune response molecules both in vitro and in vivo. As a consequence, these polymorphisms in the regulatory parts of the genes are probably responsible for the individual variation in the levels of several cytokines, a phenomenon described as “high” versus “low” producers of particular cytokines.

3. Possibilities for assessing DNA polymorphisms in the immune response genes

In general, primary identification of DNA (gene) polymorphism is based on comparison of the homologous DNA sequences among individuals of the same species. Once DNA polymorphism is identified, it may be determined (“genotyped”) in the population sample by tens of genotyping techniques which exist in hundreds of modifications. Genotyping techniques have been repeatedly reviewed. We, therefore, limit ourselves to examples of the “classical” genotyping techniques characterised by relatively low throughput. In contrast, novel “high-throughput” genotyping technologies have dramatically improved the possibilities of genetic epidemiology – they allow parallel genotyping of thousands of variants. Accordingly, the first whole genome association studies on the genetic susceptibility to common diseases have already been published. Finally, DNA sequencing as a tool for identifying novel variants and prospectively an alternative to whole genome genotyping is mentioned here.

3.1 Examples of “classical” genotyping techniques

PCR-SSP (Polymerase chain reaction with sequence specific primers) is based on allele discrimination during the PCR amplification – under stringent conditions the sequence specific primers targeted to the particular allele will produce the product only for this allele and not amplify closely related alleles non-specifically. In PCR-SSO (Polymerase chain reaction with sequence specific oligonucleotides), the DNA amplified by PCR is transferred to suitable surface (dot-blot). Labelled sequence specific probes are hybridised to DNA and then detected only in the case of total complementarity. The probes are visualised by colour reaction or fluorescence. Modification of PCR-SSO with specific immobilised probes is called reverse (r)SSO technique. PCR-RFLP (Polymerase chain reaction with restriction fragment length polymorphism)
is one of the earliest methods to detect SNPs. This technique uses a number of different restriction endonucleases and their high affinity to unique and specific restriction sites. By performing digestion on a PCR product containing SNP position and determining fragment lengths through a gel assay, it is possible to ascertain whether or not the enzymes cut the expected restriction sites. Taq DNA polymerase’s 5’-nuclease activity is used in the “5’-nuclease assay” (TaqMan) for SNP genotyping. The assay requires forward and reverse PCR primers that will amplify a region that includes the SNP polymorphic site. Allele discrimination is achieved using one or two allele-specific probes hybridizing to the SNP polymorphic site that have a fluorophore linked to their 5’ end and a quencher molecule linked to their 3’ end. In the case of total complementarity of the allele-specific probe to the SNP allele, it will bind to the target DNA strand and then be degraded by the 5’-nuclease activity of the Taq polymerase as it extends the DNA from the PCR primers. The degradation of the probe results in the separation of the fluorophore from the quencher molecule, generating a detectable signal.

3.2 “High throughput” genotyping including “whole genome” microarrays
3.2.1 Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF)
Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) is an analytical method of mass spectrometry that carries out analysis of the synthetic oligonucleotides including modified sequences. The technique is based on the ionization of oligonucleotides using matrix and follow-up detection based on transit time measurement of ionized molecules through flight tube and determination appropriate molecular weight. Based on the interpretation of the mass spectrum, it is possible to establish the exact synthesized sequence, the molecular weight of appropriate contamination and practice oligonucleotides sequencing to a determined degree of precision.

3.2.2 SNP microarrays
In high density oligonucleotide SNP arrays, hundreds of thousands of probes are arrayed on a small chip, allowing for a large number of SNPs to be investigated simultaneously. As SNP alleles only differ in one nucleotide and because it is difficult to achieve optimal hybridization conditions for all probes on the array, the target DNA has the potential to hybrdize to mismatched probes having the SNP site in several different locations as well as containing mismatches to the SNP allele. By comparing the differential amount of hybridization of the target DNA to each of these redundant probes, it is possible to determine specific homozygous and heterozygous alleles. Although, oligonucleotide microarrays have a comparatively lower specificity and sensitivity, the scale of SNPs that can be investigated is a major benefit. The most common and most used techniques in microarray technologies are produced under the following commercial marks: Illumina and Affymetrix.

Illumina – this technology works on the principle of so called “BeadArray”. It is based on 3-micron silica beads that self assemble in microwells on either of two substrates: fiber optic bundles or planar silica slides. Each bead is covered with hundreds of thousands of copies of a specific oligonucleotide that act as the capture sequences in one of Illumina’s assay (http://www.illumina.com/).

Affymetrix – this method is able to detect thousands of SNPs using hundreds of nanograms of sample DNA based on measuring the hybridization power of oligonucleotides with appropriate particular sequence variants, as controls use oligonucleotides with a purposeful mistake inside (http://www.affymetrix.com/). With the 500k GeneChip it is theoretically possible to provide genotypes for 500000 SNPs, on two separate chips.

3.3 DNA sequencing
DNA sequencing techniques determine the order of the nucleotide bases in the investigated template DNA. In terms of genotyping they produce the most exact information on the investigated DNA sequence as they are not limited to the sites of “known” variants but may identify novel variants. At present, the chain termination method (dideoxy sequencing or Sanger method) which exists in various modifications is the most commonly used DNA sequencing technique. The reaction mix contains DNA template, primer, four dNTPs and four ddNTPs labelled by different fluorescent dyes and its products, detectable on capillary electrophoresis, are several chains with primer at the 5’-end and mixture specific ddNTP on the other side. Recently, several platforms for “ultra-high throughput” sequencing based on revolutionary technologies have been introduced (e.g. “SOLID” System by Applied Biosystems company). These approaches may allow sequencing whole genomes in the near future with incomparably lower costs and time demands than classical sequencing techniques.

4. Implication of immune response genes (IRG) in the susceptibility to disease
Classical genetics divides diseases according to the genetic component involved into: 1) monogenic diseases, where the single gene variant (mutation) of the gene directly causes the disease, 2) complex (multigenic) diseases, where the interaction of the genetically susceptible person with environmental factors leads to the disease development, and 3) diseases without genetic component which develops independently of the genome variation.

Variation in immune response genes are implicated in both monogenic and complex diseases. Knowledge of the molecular mechanisms associated with monogenic diseases has been rapidly growing together with developments in the area of molecular genetics. The causative genetic variants have been identified for a large number of monogenic diseases, e.g. for primary immunodeficiencies. In general, these variants cause critical changes in the structure and/or expression of the IRG but their frequencies in the general population are very low (often not considered as “polymorphism” according to the arbitrary “one percent” definition).
Table 1. Genome-wide association studies for identification of susceptibility genetic variants for complex diseases: examples.

<table>
<thead>
<tr>
<th>REFERENCE</th>
<th>DISEASE</th>
<th>CHROMOSOME/GENE</th>
<th>SNP</th>
<th>ASSOCIATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim KM et al.(^{53})</td>
<td>Chronic obstructive pulmonary disease (COPD)</td>
<td>COL4A3</td>
<td>H451R</td>
<td>COL4A3 contributes to the genetic susceptibility to COPD</td>
</tr>
<tr>
<td>Chapuis J et al.(^{54})</td>
<td>Alzheimer’s disease</td>
<td>GAB2</td>
<td>rs2373115</td>
<td>risk of developing Alzheimer’s disease</td>
</tr>
<tr>
<td>Sun J et al.(^{55})</td>
<td>Prostate cancer (PC)</td>
<td>17q12 17q24.3</td>
<td>rs4430796  rs1859962</td>
<td>earlier age at diagnosis of the PC</td>
</tr>
<tr>
<td>Liu YJ et al.(^{56})</td>
<td>Obesity</td>
<td>CTNNBL1</td>
<td>rs6013029</td>
<td>minor allele confers an average increase in BMI and fat mass</td>
</tr>
<tr>
<td>Prescott NJ et al.(^{57})</td>
<td>Crohn’s disease</td>
<td>ATG16L1</td>
<td>rs2241880</td>
<td>Ulcerative colitis</td>
</tr>
<tr>
<td>Oishi T et al.(^{58})</td>
<td>Systemic lupus erythematosus (SLE)</td>
<td>ITPR3</td>
<td>rs3748079</td>
<td>pathogenesis of SLE</td>
</tr>
<tr>
<td>Samani NJ et al.(^{59})</td>
<td>Coronary artery disease (CAD)</td>
<td>9p21 6q25.1   2q36.3 1p13.3 1q41 10q11.21 15q22.33</td>
<td>rs133049  rs6922269 rs2943634  rs599839 rs17465637 rs501120 rs17228212</td>
<td>affect the risk of development of CAD</td>
</tr>
<tr>
<td>McPherson R et al.(^{60})</td>
<td>Coronary heart disease (CHD)</td>
<td>9p21</td>
<td>rs10757274  rs2383206</td>
<td>increased risk of CHD by the homozygotes for the risk allele</td>
</tr>
<tr>
<td>Wellcome Trust Case Control Consortium(^{61})</td>
<td>Bipolar disorder (BD)  Coronary artery disease (CAD)  Crohn’s disease (CD)</td>
<td>16p12 9p21 1p31 2q37 3p21 5p13 5q33 10q21 10q24 16q12 8p11 1p13 6 7q32</td>
<td>rs420259  rs133049  rs11805303  rs10210302  rs9858542  rs17234657  rs1000113  rs10761659  rs10883365  rs17221417  rs2542151  rs6679677  rs6457617  rs11761231</td>
<td>BD pathogenesis  CAD pathogenesis  CD susceptibility  strong additive effect in females preventing autoimmunity  strongest association signal to DT2</td>
</tr>
</tbody>
</table>
Table 2. Selected studies investigating possible association of candidate genetic variants with susceptibility to sarcoidosis.

<table>
<thead>
<tr>
<th>REFERENCE</th>
<th>GENE</th>
<th>SNP/VARIANT</th>
<th>SUSCEPTIBILITY/CLINICAL COURSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navratilova Z et al.62</td>
<td>MCP-1 - 2518</td>
<td>A/G</td>
<td>Löfgren’s syndrome</td>
</tr>
<tr>
<td>Mauer A et al.63</td>
<td>IL6-174</td>
<td>G/C</td>
<td>Sarcoïdosis</td>
</tr>
<tr>
<td>Valentonyte R et al.37</td>
<td>BTNL2 (rs2076530)</td>
<td>G/A</td>
<td>Sarcoïdosis</td>
</tr>
<tr>
<td>Zorzetto M et al.42</td>
<td>CR1 C5507G</td>
<td>C/G</td>
<td>Sarcoïdosis</td>
</tr>
<tr>
<td>Kruit A et al.64</td>
<td>CMA1</td>
<td>C/T</td>
<td>vital inspiratory capacity (functional outcome of pulmonary sarcoidosis)</td>
</tr>
<tr>
<td>Mrazek F et al.35</td>
<td>TNFα LTα</td>
<td>G/A/A/G</td>
<td>Löfgren’s syndrome</td>
</tr>
<tr>
<td>Spagnolo P et al.45</td>
<td>CCR2</td>
<td>T/G A/G T/C G/C T/A</td>
<td>Löfgren’s syndrome</td>
</tr>
<tr>
<td>Swider C et al.14</td>
<td>TNFα HLA</td>
<td>-308 DRB1</td>
<td>Löfgren’s syndrome</td>
</tr>
<tr>
<td>Rutherford RM et al.33</td>
<td>HLA</td>
<td>DR2</td>
<td>chronic sarcoidosis</td>
</tr>
<tr>
<td>Sato H et al.32</td>
<td>HLA</td>
<td>DQB1</td>
<td>Löfgren’s syndrome and progression of the disease</td>
</tr>
<tr>
<td>Rybicki BA et al.38</td>
<td>BTNL2</td>
<td>rs2076530</td>
<td>Sarcoïdosis</td>
</tr>
<tr>
<td>Spagnolo P et al.31</td>
<td>BTNL2 HLA</td>
<td>rs2076530 DRB1</td>
<td>non- Löfgren sarcoidosis</td>
</tr>
</tbody>
</table>

It has been shown in epidemiological studies (e.g. studies on twins) that the great majority of diseases with significant impact on total morbidity and mortality in the developed countries (e.g. atherosclerosis and its complications, allergy / asthma, cancer, autoimmunity disorders) are complex disorders with various genetic factor contributions. Since molecules encoded by IRG (e.g. mediators, receptors, MHC complex) are substantially implicated in the development of these diseases, IRG functional polymorphisms are plausible candidates for the participation in the susceptibility to these diseases.

Nevertheless, compared with the tight relationship between cause (gene variant) and consequence (disease) for monogenic disorders, the genetics of complex diseases may be understood as rather probabilistic, often with participation of complicated gene - environment and gene-gene interactions. Further paragraphs aim at a brief description of the main approaches used for identification of genetic variants associated with these diseases. At present, how a particular genetic variant contributes to a particular disease with sufficient evidence should comply with the rules / recommendations developed for this type of genetic research (Fig. 3)26. In general, exact definition and recruitment of disease phenotype (“phenotype homogeneity”), correct genotyping and statistical analysis, consideration of population background (stratification in the selection of control subjects) and linkage disequilibrium would all be necessary for conducting such genetic studies. Importantly, validation of the primary data by replication in independent studies (groups) has appeared in the last decade to be necessary for confirmation of the role of genetic variants in diseases.

4.1 Candidate gene approach versus genome wide studies in susceptibility to complex diseases

Recently, rapid development (examples mentioned in the section on high throughput genotyping techniques) over the last few years have provided the possibility of performing the first genome wide genetic studies for detection of the susceptibility loci for complex diseases.
Genome wide studies have completely revolutionarised the area of the genetics of complex diseases which was previously limited to the usage of “candidate” gene approach.

Genome wide studies apply an “hypothesis free” approach; they test genetic markers which sufficiently cover the total variation of the genome. For this purpose, an extremely large number of variants (up to half a million SNPs) is genotyped in a group of patients with a particular disease and their distributions are then compared with those found in a healthy control population (see Table 1 for examples). Application of the candidate gene approach implies a knowledge of the principal molecular mechanisms of the disease. In this regard, candidate genes may be defined as those which “mediate” the disease in the broadest sense. Accordingly, variants of these genes (preferentially of known functionality) are selected and investigated in patients and control subjects (or first grade relatives).

4.2. Design of studies on genetic susceptibility to complex diseases

4.2.1 Linkage analysis

Linkage analysis is a method for determining if there is significant evidence for co-segregation of alleles of a marker with alleles at a hypothetical disease locus. Accordingly, this approach is based on analysis of the segregation of investigated genetic markers within families. Two basic types of linkage analysis are generally recognized. Parametric (classical) linkage analysis may be used especially in simple (monogenic) Mendelian diseases, when the model of inheritance is correctly specified. Non-parametric linkage analysis (e.g. affected sibling pair method using genome scans) is based on analysing whether alleles of particular genetic marker are shared between siblings affected by the diseases more often than would be expected by their random segregation. By comparison with case – control design (see below), linkage analysis has limited power to detect disease susceptibility alleles conferring moderate/mild risk to the disease. Close attention to the evidence for a genetic factor and careful selection of the disease phenotype may, however, increase the power of this approach.

4.2.2 Case – control association studies

Association studies are the most frequently used approach to identifying susceptibility genes for complex diseases at present. They are based on comparison of the frequencies of “candidate” gene variant between patients and healthy controls. Population-based association studies are characterised by selecting control subjects from ethnically matched population, whereas family-based association studies use the members (usually parents) of the affected relative. Genetic association studies are more powerful than linkage analysis in detecting genes with small effects in complex diseases. On the other hand, these studies are more prone to false positive results. Most of these studies focus directly on a single gene and frequently look directly at functionally significant polymorphisms.

4.3 Bioinformatics and biostatistics in the genetic research

Current knowledge and approaches to genetic susceptibility to complex diseases is, like global biomedical research, unimaginable without using publicly available...
Polymorphisms of the immune response genes: selected biological, methodical and medical aspects

4.4 Diseases and genetic susceptibility variants – model examples

The term “genetic susceptibility to complex disease” can be divided into levels with potentially important clinical / therapeutical consequences. Gene polymorphisms that predispose to the development disease *per se* (“disease-susceptibility variants”) are overrepresented in the whole patient population. “Disease-modifying variants” is a term reserved for genetic polymorphisms associated with particular phenotype or affecting the clinical course of the disease. Disease-modifying variants may be associated with the prognosis of the disease and their prospective analysis may eventually be used in the clinical protocols. Finally, a large number of genetic variants are associated with drug metabolism and, therefore, may have important consequences for disease treatment. In the following section we outline current knowledge on genetic susceptibility to the selected complex diseases, sarcoidosis and coronary artery disease which have been investigated also by the authors of this review.

4.4.1 Genetics of sarcoidosis

Sarcoidosis is a chronic granulomatous inflammatory disease with unknown etiology most frequently affecting the lung. Sarcoidosis likely develops from complex interactions between environmental agents and the alleles of several genes (multigenic susceptibility). Despite the unknown environmental causes of sarcoidosis, there is strong evidence for a genetic component based on earlier epidemiological studies (e.g. differences in the incidence and “phenotype” of the disease among different ethnic groups and overrepresentation among the relatives). Since the first reports describing sarcoidosis in siblings, main histocompatibility complex (HLA) and a huge number of further candidate immune response genes (IRG) have been tested for possible association with sarcoidosis susceptibility. Candidate variants have been chosen from IRG encoding for 1) molecules implicated in recognition, processing and presentation of the antigens, and 2) mediators controlling for T cell functions, macrophage activation and granulomatous inflammation. Association studies have been motivated by the hope that identifying alleles that affect risk and phenotype of sarcoidosis will help in understanding disease etiology.

Several examples of the reports on genetic susceptibility to sarcoidosis are shown in Table 2. It can be seen from the table that there is most focus on the HLA genes. Thus, one of the most consistent findings throughout the literature on sarcoidosis genetics is association of the HLA alleles DRB1*03 (particularly DRB1*0301 – DR17) and/or closely linked DQB1*02 (DQ*0201) with better prognosis, namely its specific presentation called Löfgren’s syndrome. Importantly, genes for other crucial mediators of granulomatous inflammation – tumour necrosis factor (TNFα) and lymphotixin (LTα) are located within the HLA genetic region. TNFα stimulates cytokine production, enhances expression of adhesion molecules, and databases. Of these the databases and tools in publicly accessible U.S. National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) webside are used worldwide for search/selection for candidate genes, polymorphisms, sequence alignments and current literature in the field. Another database providing a huge amount of priority data for genetic research may be found at the webside of International HapMap project (http://www.hapmap.org/). The HapMap project is particularly aimed at development and management of a public resource that helps researchers find genes associated with human disease and response to pharmaceuticals.

In addition to using public databases, research on the genetic component of diseases would not be possible without application of highly sophisticated software developed for the collection, management and interpretation of the data. Of particular importance, growing experience from genetic association studies including commonly observed discordance in results among various centres results in the formulation of the rules for this type of genetic research. Accordingly, the requirements for the application of appropriate statistical analysis in the interpretation of the data on association of genetic variants with diseases have markedly increased. Indeed, novel strategies in genetic research, namely the genome-wide approach, require the cooperation of experts in bioinformatics/biostatistics.

Fig. 3. Identification / confirmation of the association between particular genetic variant and disease susceptibility. In the initial phase, exact definition of aims considering statistical power and genetic background should be performed. Relevant recruitment of the cases / controls and sufficient homogeneity of disease phenotype are further essential prerequisites. Genetic and analytical part of this type of studies requires application of reliable genotyping and correct analysis / interpretation of the data. Phase of confirmation which involves replication of the results in further population(s) and/or functional evidence on obtained variants may be involved in the primary study.

<table>
<thead>
<tr>
<th>Genetic variant</th>
<th>Disease phenotype</th>
<th>Population background</th>
<th>Statistical power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotyping, analysis, interpretation</td>
<td>Reliable genotyping</td>
<td>Linkage disequilibrium</td>
<td></td>
</tr>
<tr>
<td>Phase of confirmation</td>
<td>Functional evidence</td>
<td>Replication</td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
acts as a costimulator of T-cell activation. It has also been shown that alveolar macrophages of patients with active sarcoidosis secrete more TNFα than those with inactive disease. In this context, allelic variants of the TNF gene have also been associated with susceptibility to sarcoidosis. Particularly, the TNF-857*T allele has been reported as a susceptibility allele, whereas the variant TNF-308*A was found to be associated with Löfgren’s syndrome and erythema nodosum, respectively. Finally, truncating splice site mutation of the BTNL2 gene located in close proximity to the HLA genes has been found to be associated with sarcoidosis, whereas the variant TNF-308*A was found to be associated with Löfgren’s syndrome and erythema nodosum, respectively. Numerous other candidate variants have been investigated (and some of them confirmed) in sarcoidosis susceptibility, e.g. genes for chemokine receptors CCR2, CCR5, Heat Shock Protein (HSP)A, Clara cell 10-kD protein or Toll-like receptor 4 (TLR4). Unfortunately, a significant proportion of the genetic associations with sarcoidosis have not been replicated in other studies; e.g. the Complement receptor 1 (CR1) – 5507 GG genotype significantly associated with sarcoidosis in an Italian population had no effect on sarcoidosis susceptibility in a Czech population.

4.4.2 Genetics of coronary artery disease – promising whole genome approach

Coronary artery disease (CAD) is one of the most common causes of death in the developed countries. CAD develops usually as a clinical manifestation of atherosclerosis which is considered currently as a process strongly connected with inflammation. Like sarcoidosis, the epidemiology of CAD earlier provided clear evidence for the genetic component of this disease. However, a plethora of external and internal factors are implicated in the susceptibility to atherosclerosis (including CAD) – e.g. lipid metabolism, diabetes, hypertension, obesity, diet, and smoking. Accordingly, the spectrum of candidate genes in CAD is extremely broad – from genes controlling inflammation (the members of the IRG group), through those implicated in lipid metabolism to the genes of the coagulation cascade. In the complex genetic base of CAD, phenotype and genetic heterogeneity, low penetrance of the susceptibility alleles, and high frequency of susceptibility alleles in the population should be considered. The literature on the genetic association with CAD is huge. However, an alarmingly high proportion of reported associations between genetic variants and CAD have not been replicated. However given the huge social impact of CAD, a major effort has been dedicated to the “whole genome” studies in CAD recently. These studies (reviewed in Table 1) have brought promising results: several genetic loci associated with CAD have repeatedly been replicated. Of these e.g. chromosomal region 9p21.3 is confirmed as a susceptibility marker for CAD.

5. Perspectives on the research and clinical applications of the knowledge on immune response genes

It should be remembered that the research findings on IRG are already being routinely used in a number of fields in clinical medicine. The immunogenetics of the HLA system are applied in clinical transplantation (donor–recipient matching) and complementary diagnostics of HLA-associated diseases. The possibilities for identification and diagnostics of monogenic diseases (e.g. primary immunodeficiencies) have been markedly increased by investigation of causative genetic variants located most frequently in the IRG. Apart from these clearly “practical” clinical applications, investigations of IRG have dramatically improved our understanding of the mechanisms of the immune system and immune response.

The research of IRG and genome research in general confronts major challenges. From the technological point of view, despite the rapid developments in high-throughput genotyping technology, the cost and time for sequencing individual human genome still prevent its use in genetics association studies. Similarly, the management and proper statistical analysis/interpretation of the “gigabites” of data obtained from large genetics studies require deeper understanding of biostatistics and bioinformatics.

As main aims of future research into genetic susceptibility to diseases we can suggest: 1) definition of the genetic component of the diseases in terms of particular genetic variants, 2) identification of novel treatment targets, 3) contribution to individualised medicine – role in prognosis of the disease and treatment (including pharmacogenomics), and 4) diagnosis of the diseases (reserved for strongly linked variants, especially in monogenic disorders). On the basis of recent data from genome-wide association studies, it is now clear that rather than to predict diseases there will be greater clinical benefit of genome research in better understanding etiology and disease processes.

ACKNOWLEDGEMENTS

This study was supported by grants IGA NR9099-4 and IGA NR9490-3.

REFERENCES


61. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007; 447:661-78.