

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 databases⁵. 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 another⁶.

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 core 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 identification⁷. 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 cases⁸. 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 reviewed⁹⁻¹¹. We, therefore, limit ourselves to examples of the “classical” genotyping techniques characterized 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 published¹²⁻¹⁴. 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-specifically¹⁵. 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 complementarity¹⁶. 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 sites¹⁷. 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^{18, 19}.

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 practise oligonucleotides sequencing to a determined degree of precision²⁰.

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 simultaneously²¹. 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 hybridize 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 alleles²¹. 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 technique²². 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 techniques²³.

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 variation²⁴.

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 immunodeficiencies²⁵. 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.

REFERENCE	DISEASE	CHROMOSOME/ GENE	SNP	ASSOCIATION
Kim KM et al. ⁵³	Chronic obstructive pulmonary disease (COPD)	COL4A3	H451R	COL4A3 contributes to the genetic susceptibility to COPD
Chapuis J et al. ⁵⁴	Alzheimer's disease	GAB2	rs2373115	risk of developing Alzheimer's disease
Sun J et al. ⁵⁵	Prostate cancer (PC)	17q12 17q24.3	rs4430796 rs1859962	earlier age at diagnosis of the PC
Liu YJ et al. ⁵⁶	Obesity	CTNBL1	rs6013029	minor allele confers an average increase in BMI and fat mass
Prescott NJ et al. ⁵⁷	Crohn's disease	ATG16L1	rs2241880	Ulcerative colitis
Oishi T et al. ⁵⁸	Systemic lupus erythematosus (SLE)	ITPR3	rs3748079	pathogenesis of SLE
Samani NJ et al. ⁵⁹	Coronary artery disease (CAD)	9p21.3 6q25.1 2q36.3 1p13.3 1q41 10q11.21 15q22.33	rs1333049 rs6922269 rs2943634 rs599839 rs17465637 rs501120 rs17228212	affect the risk of development of CAD
McPherson R et al. ⁶⁰	Coronary heart disease (CHD)	9p21	rs10757274 rs2383206	increased risk of CHD by the homozygotes for the risk allele
Wellcome Trust Case Control Consortium ⁶¹	Bipolar disorder (BD)	16p12	rs420259	BD pathogenesis
	Coronary artery disease (CAD)	9p21	rs1333049	CAD pathogenesis
	Crohn's disease (CD)	1p31 2q37 3p21 5p13 5q33 10q21 10q24 16q12 8p11 1p13 6 7q32	rs11805303 rs10210302 rs9858542 rs17234657 rs1000113 rs10761659 rs10883365 rs17221417 rs2542151 rs6679677 rs6457617 rs11761231	CD susceptibility
	Rheumatoid arthritis			strong additive effect in females preventing autoimmunity
	Diabetes type 1	10p15 1p13 6 12q13 12q24 16p13 4q27 12p13	rs2104286 rs6679677 rs9272346 rs11171739 rs17696736 rs12708716 rs6534347 rs3764021	
	Diabetes type 2	6p22 10q25	rs9465871 rs4506565	strongest association signal to DT2
		16q12	rs9939609	

Table 2. Selected studies investigating possible association of candidate genetic variants with susceptibility to sarcoidosis.

REFERENCE	GENE	SNP/VARIANT	SUSCEPTIBILITY/CLINICAL COURSE
Navratilova Z et al. ⁶²	MCP-1 - 2518	A/G	Löfgren's syndrome
Maver A et al. ⁶³	IL6-174	G/C	Sarcoidosis
Valentonyte R et al. ³⁷	BTNL2 (rs2076530)	G/A	Sarcoidosis
Zorzetto M et al. ⁴⁵	CR1 C5507G	C/G	Sarcoidosis
Kruit A et al. ⁶⁴	CMA1	C/T	vital inspiratory capacity (functional outcome of pulmonary sarcoidosis)
Mrazek F et al. ³⁵	TNF α LT α	G/A A/G	Löfgren's syndrome
Spagnolo P et al. ⁶⁵	CCR2	T/G A/G T/C G/C T/A	Löfgren's syndrome
Swider C et al. ³⁴	TNF α HLA	-308 DRB1	Löfgren's syndrome
Rutherford RM et al. ³³	HLA	DR2	chronic sarcoidosis
Sato H et al. ³²	HLA	DQB1	Löfgren's syndrome and progression of the disease
Rybicki BA et al. ³⁸	BTNL2	rs2076530	Sarcoidosis
Spagnolo P et al. ³¹	BTNL2 HLA	rs2076530 DRB1	non- Löfgren sarcoidosis

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.

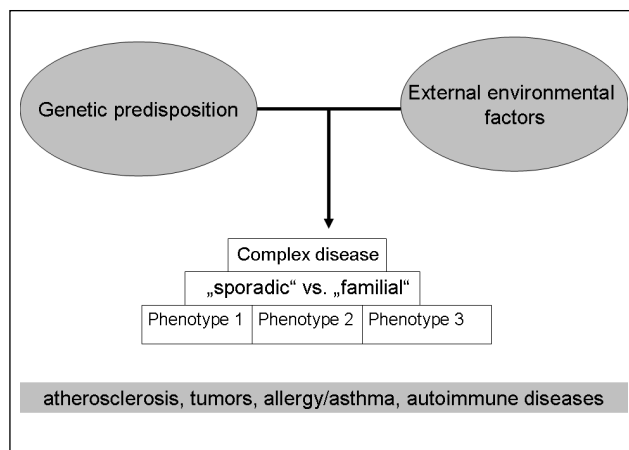


Fig. 1. „Classical” model of the pathogenesis of the complex disease. Both genetics and environmental factors interact in various proportions in the development of the diseases. In the majority of complex diseases, “familial” and “sporadic” cases may be distinguished based on epidemiological definition. Furthermore, distinct disease phenotypes may be observed within the diagnosis of the complex disease, often characterised by different prognosis.

Genome wide studies have completely revolutionised 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 locus²⁷. 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

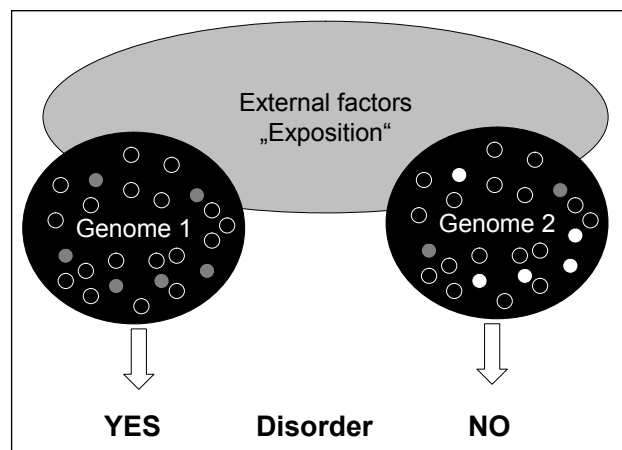


Fig. 2. Theoretical scheme of the role of the genetic component in the development of complex disease. Two individuals represented by their genomes were exposed to identical external (environmental) factors. Genome 1 carrying several risk genetic variants (grey circles) promote development of the disease. Contrary, protective (neutral) variants (white circles) on the same genetic loci in genome 2 confer non-development of the disease. The great majority of the variant sites (empty circles) in the genome has no effect on the susceptibility to disease.

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

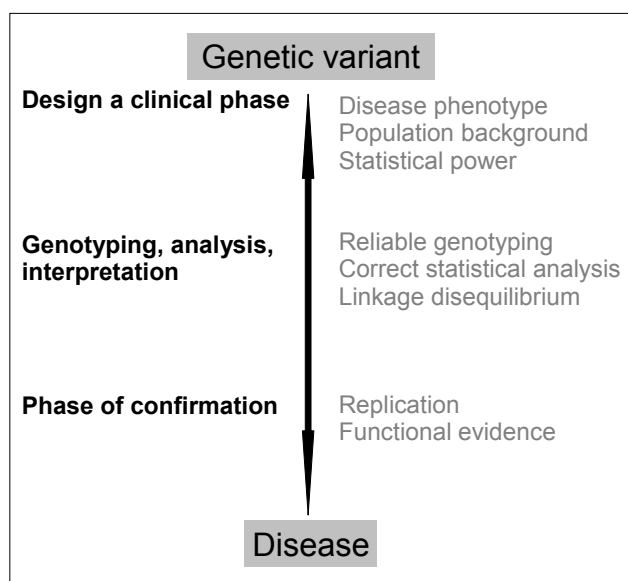


Fig. 3. Identification / confirmation of the association between particular genetic variant and disease susceptibility. In the initial phase, exact definition of aims considering statistical power and genetic background should be performed. Relevant recruitment of the cases / controls and sufficient 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.

databases. Of these the databases and tools in publicly accessible U.S. National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>) website 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 website 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^{28, 29}. 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.

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 lung³⁰. 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^{31–34}. 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^{32–35}. Importantly, genes for other crucial mediators of granulomatous inflammation – tumour necrosis factor (TNF) α and lymphotoxin (LT) α are located within the HLA genetic region. TNF α stimulates cytokine production, enhances expression of adhesion molecules, and

acts as a costimulator of T-cell activation. It has also been shown that alveolar macrophages of patients with active sarcoidosis secrete more TNF α than those with inactive disease³⁰. 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 allele³⁶, whereas the variant TNF-308*A was found to be associated with Löfgren's syndrome and erythema nodosum, respectively^{34, 35}. 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^{37, 38, 31}. Numerous other candidate variants have been investigated (and some of them confirmed) in sarcoidosis susceptibility, e.g. genes for chemokine receptors CCR2, CCR5^{39, 40}, Heat Shock Protein (HSP)A^{41, 42}, Clara cell 10-kD protein⁴³ or Toll-like receptor 4 (TLR4)⁴⁴. 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 population⁴⁵ had no effect on sarcoidosis susceptibility in a Czech population⁴⁶.

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 considered⁴⁷. 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 replicated⁴⁸. 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 CAD⁴⁹⁻⁵¹.

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 "gigabytes" 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 processes⁵².

ACKNOWLEDGEMENTS

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

REFERENCES

1. Krejsek J, Kopecký O. Klinická imunologie. Nucleus, Hradec Králové; 2004.
2. Bochud PY, Bochud M, Telenti A, Calandra T. Innate immunogenetics: a tool for exploring new frontiers of host defence. *Lancet Infect Dis.* 2007; 7: 531-42.
3. Schwartz DA. The genetics of innate immunity. *Chest* 2002; 121 (3 Suppl): 62S-68S.
4. Rihet P. Innate immunity genes as candidate genes: searching for relevant natural polymorphisms in databases and assessing family-based association of polymorphisms with human diseases. *Methods Mol Biol.* 2008; 415: 17-48.
5. Lee JE. High-throughput genotyping. *Forum Nutr.* 2007; 60: 97-101.
6. Hoffmann SC, Stanley EM, Cox ED, DiMercurio BS, Kozioł DE, Harlan DM et al. Ethnicity greatly influences cytokine gene polymorphism distribution. *Am J Transplant.* 2002; 2: 560-7.
7. Cheng BW, Chen GD, Zhang HJ. A study of forensic individual identification used STR locus with silver staining and multiplex PCR methods. *Yi Chuan.* 2002; 24: 15-8.
- Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz RL, Wallace RB. Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proceedings of the National Academy of Science of the United States of America* 1983; 80: 278-82.

8. Lee CI, Leong SH, Png AE, Choo KW, Syn C, Lim DT et al. An isothermal method for whole genome amplification of fresh and degraded DNA for comparative genomic hybridization, genotyping and mutation detection. *DNA Res.* 2006; 13: 77-88.
9. Maresso K, Broeckel U. Genotyping platforms for mass-throughput genotyping with SNPs, including human genome-wide scans. *Adv Genet.* 2008; 60: 107-39.
10. Bier FF, von Nickisch-Rosenegk M, Ehrentreich-Förster E, Reiss E, Henkel J, Strehlow R et al. DNA microarrays. *Adv Biochem Eng Biotechnol.* 2008; 109: 433-53.
11. Kim S, Misra A. SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng.* 2007; 9: 289-320.
12. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; 316: 1331-6.
13. Yang HH, Hu N, Taylor PR, Lee MP. Whole genome-wide association study using affymetrix SNP chip: a two-stage sequential selection method to identify genes that increase the risk of developing complex diseases. *Methods Mol Med.* 2008; 141: 23-35.
14. Loeuillet C, Deutsch S, Ciuffi A, Robyr D, Taffé P, Muñoz M et al. In vitro whole-genome analysis identifies a susceptibility locus for HIV-1. *PLoS Biol.* 2008; 6: e32.
15. Kwok PY. Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet.* 2001; 2: 235-58.
16. Angelini G, de Preval C, Gorski J, Mach B. High-resolution analysis of the human HLA-DR polymorphism by hybridization with sequence-specific oligonucleotide probes. *Proceedings of the National Academy of Science of the United States of America* 1986; 83: 4489-93.
17. Ota M, Fukushima H, Kulski JK, Inoko H. Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. *Nat Protoc.* 2007; 2: 2857-64.
18. McGuigan FE, Ralston SH. Single nucleotide polymorphism detection: allelic discrimination using TaqMan. *Psychiatric Genetics* 2002; 12: 133-136.
19. Syvänen AC. Accessing genetic variation : genotyping single nucleotide polymorphisms. *Nature Reviews Genetics* 2001; 2: 930 - 942.
20. Ragoussis J, Elvidge GP, Kaur K, Colella S. Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. *PLoS Genet.* 2006; 2: e100.
21. Rapley R, Harbron S. Molecular analysis and genome discovery. Wiley; 1 edition May 26 2004. Rapley R, Harbron S, editors. Wiley; 2004.
22. França LT, Carrilho E, Kist TB. A review of DNA sequencing techniques. *Q Rev Biophys.* 2002; 35: 169-200.
23. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods.* 2008; 5: 16-8.
24. Ropers HH. New perspectives for the elucidation of genetic disorders. *Am J Hum Genet.* 2007; 81: 199-207.
25. Maródi L, Notarangelo LD. Immunological and genetic bases of new primary immunodeficiencies. *Nat Rev Immunol.* 2007; 7: 851-61.
26. Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005; 2: e124.
27. Tang WC, Yap MK, Yip SP. A review of current approaches to identifying human genes involved in myopia. *Clin Exp Optom.* 2008; 91: 4-22.
28. Cordell HJ, Clayton DG. Genetic association studies. *Lancet.* 2005; 366: 1121-31.
29. NCI-NHGRI Working Group on Replication in Association Studies, Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G et al. Replicating genotype-phenotype associations. *Nature.* 2007; 447: 655-60.
30. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med.* 2007; 357: 2153-65.
31. Spagnolo P, Sato H, Grutters JC, Renzoni EA, Marshall SE, Ruven HJ et al. Analysis of BTNL2 genetic polymorphisms in British and Dutch patients with sarcoidosis. *Tissue Antigens* 2007; 70: 219-27.
32. Sato H, Grutters JC, Pantelidis P, Mizzon AN, Ahmad T, Van Houte AJ et al. HLA-DQB1*0201: a marker for good prognosis in British and Dutch patients with sarcoidosis. *Am J Respir Cell Mol Biol.* 2002; 27: 406-12.
33. Rutherford RM, Brutsche MH, Kearns M, Bourke M, Stevens F, Gilmartin JJ. (2004) HLA-DR2 predicts susceptibility and disease chronicity in Irish sarcoidosis patients. *Sarcoidosis Vasc Diffuse Lung Dis.* 2004; 21: 191-8.
34. Swider C, Schnittger L, Bogunia-Kubik K, Gerdes J, Flad H, Lange A et al. TNF-alpha and HLA-DR genotyping as potential prognostic markers in pulmonary sarcoidosis. *Eur Cytokine Netw.* 1999; 10: 143-6.
35. Mrazek F, Holla LI, Hutyrova B, Znojil V, Vasku A, Kolek V et al. Association of tumour necrosis factor-alpha, lymphotoxin-alpha and HLA-DRB1 gene polymorphisms with Löfgren's syndrome in Czech patients with sarcoidosis. *Tissue Antigens* 2005; 65: 163-71.
36. Grutters JC, Sato H, Pantelidis P, Lagan AL, McGrath DS, Lammers JW et al. Increased frequency of the uncommon tumor necrosis factor -857T allele in British and Dutch patients with sarcoidosis. *Am J Respir Crit Care Med.* 2002; 165: 1119-24.
37. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A et al. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet.* 2005; 37: 357-64.
38. Rybicki BA, Walewski JL, Maliarik MJ, Kian H, Iannuzzi MC; ACCESS Research Group. The BTNL2 gene and sarcoidosis susceptibility in African Americans and Whites. *Am J Hum Genet.* 2005; 77: 491-9.
39. Petrek M, Drábek J, Kolek V, Zlámál J, Welsh KI, Bunce M et al. CC chemokine receptor gene polymorphisms in Czech patients with pulmonary sarcoidosis. *Am J Respir Crit Care Med.* 2000; 162 (3 Pt 1): 1000-3.
40. Maliarik MJ, Rybicki BA, Malvitz E, Scheffer RG, Major M, Popovich J Jr et al. Angiotensin-converting enzyme gene polymorphism and risk of sarcoidosis. *Am J Respir Crit Care Med.* 1998; 158: 1566-70.
41. Bogunia-Kubik K, Koscińska K, Suchnicki K, Lange A. HSP70-hom gene single nucleotide (+2763 G/A and 12437 C/T) polymorphisms in sarcoidosis. *Int J Immunogenet.* 2006; 33: 135-40.
42. Ishihara M, Ohno S, Ishida T, Mizuki N, Ando H, Naruse T et al. Genetic polymorphisms of the TNFB and HSP70 genes located in the human major histocompatibility complex in sarcoidosis. *Tissue Antigens* 1995; 46: 59-62.
43. Ohchi T, Shijubo N, Kawabata I, Ichimiya S, Inomata S, Yamaguchi A et al. Polymorphisms of Clara cell 10-kD protein gene of sarcoidosis. *Am J Respir Crit Care Med.* 2004; 169: 180-6.
44. Iannuzzi MC, Rybicki BA. Genetics of Sarcoidosis. *Proceedings of the American Thoracic Society.* 2007; Vol 4. p 108-116.
45. Zorzetto M, Bombieri C, Ferrarotti I, Medaglia S, Agostini C, Tinelli C et al. Complement receptor 1 gene polymorphisms in sarcoidosis. *Am J Respir Cell Mol Biol.* 2002; 27: 17-23.
46. Mrazek F, Kvezereli M, Garr E, Kubistova Z, Kriegova E, Fillerova R et al. Complement receptor 1 single nucleotide polymorphisms in Czech and Dutch patients with sarcoidosis. *Tissue Antigens* 2008; 71: 77-80.
47. Robin NH, Tabereaux PB, Benza R, Korf BR. Genetic testing in cardiovascular disease. *J Am Coll Cardiol.* 2007; 50: 727-37.
48. Hamsten A, Eriksson P. Identifying the susceptibility genes for coronary artery disease: from hyperbole through doubt to cautious optimism. *J Intern Med.* 2008; 263: 538-52.
49. Schunkert H, Götz A, Braund P, McGinnis R, Tregouet DA, Mangino M et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation* 2008; 117: 1675-84.
50. Helgadóttir A, Thorleifsson G, Manolescu A, Gretarsdóttir S, Blondal T, Jonasdóttir A et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007; 316: 1491-3.
51. Hinohara K, Nakajima T, Takahashi M, Hohda S, Sasaoka T, Nakahara K et al. Replication of the association between a chromosome 9p21 polymorphism and coronary artery disease in Japanese and Korean populations. *J Hum Genet.* 2008; 53: 357-9.
52. Lango H, Weedon MN. What will whole genome searches for susceptibility genes for common complex disease offer to clinical practice? *J Intern Med.* 2008; 263: 16-27.

53. Kim KM, Park SH, Kim JS, Lee WK, Cha SI, Kim CH et al. Polymorphisms in the type IV collagen alpha3 gene and the risk of COPD. *Eur Respir J*. 2008; 32: 35-41.
54. Chapuis J, Hannequin D, Pasquier F, Benthams P, Brice A, Leber I et al. Association study of the GAB2 gene with the risk of developing Alzheimer's disease. *Neurobiol Dis*. 2008; 30: 103-6.
55. Sun J, Purcell L, Gao Z, Isaacs SD, Wiley KE, Hsu FC et al. Association between sequence variants at 17q12 and 17q24.3 and prostate cancer risk in European and African Americans. *Prostate* 2008; 68: 691-7.
56. Liu YJ, Liu XG, Wang L, Dina C, Yan H, Liu JF et al. Genome-wide association scans identified CTNBL1 as a novel gene for obesity. *Hum Mol Genet*. 2008; 17: 1803-13.
57. Prescott NJ, Fisher SA, Franke A, Hampe J, Onnie CM, Soars D et al. A nonsynonymous SNP in ATG16L1 predisposes to ileal Crohn's disease and is independent of CARD15 and IBD5. *Gastroenterology* 2007; 132:1665-71.
58. Oishi T, Iida A, Otsubo S, Kamatani Y, Usami M, Takei T et al. A functional SNP in the NKX2.5-binding site of ITPR3 promoter is associated with susceptibility to systemic lupus erythematosus in Japanese population. *J Hum Genet*. 2008; 53: 151-62.
59. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007; 357: 443-53.
60. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR et al. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007; 316: 1488-91.
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.
62. Navratilova Z, Mrazek F, Kriegova E, Huttyrova B, Kolek V, du Bois RM et al. The MCP-1-2518 (A to G) single nucleotide polymorphism in Czech patients with pulmonary sarcoidosis: association with Löfgren's syndrome. *Sarcoidosis Vasc Diffuse Lung Dis*. 2007; 24: 33-8.
63. Maver A, Medica I, Salobir B, Sabovic M, Tercelj M, Peterlin B. Polymorphisms in genes coding for mediators in the interleukin cascade and their effect on susceptibility to sarcoidosis in the Slovenian population. *Int J Mol Med*. 2007; 20: 385-90.
64. Kruit A, Grutters JC, Ruven HJ, Sato H, Izumi T, Nagai S et al. Chymase gene (CMA1) polymorphisms in Dutch and Japanese sarcoidosis patients. *Respiration*. 2006; 73: 623-33.
65. Spagnolo P, Renzoni EA, Wells AU, Sato H, Grutters JC, Sestini P et al. C-C chemokine receptor 2 and sarcoidosis: association with Löfgren's syndrome. *Am J Respir Crit Care Med*. 2003; 168: 1162-6.