Coll. Antropol. **28** (2004) 2: 483–507
UDC 577.2:616-084
Editorial/Perspective

From Genomic Advances to Public Health Benefits: The Unbearable Lightness of Being Stuck

Igor Rudan[1,2] and Pavao Rudan[3]

[1] Department of Medical Statistics, Epidemiology and Medical Informatics, School of Public Health »Andrija Štampar«, School of Medicine, University of Zagreb, Croatia

[2] Department of Public Health Sciences, The University of Edinburgh Medical School, Edinburgh, Scotland, UK

[3] Institute for Anthropological Research, Zagreb, Croatia

A B S T R A C T

Genetic determinants of common human diseases are still poorly understood. Due to large investments, many small successes have been made and the research field is rapidly expanding. However, genetic susceptibility variants showing repeatable associations with common diseases are usually of small effect. They are therefore unlikely to individually explain substantial share of disease burden in any community or provide new insights into disease pathogenesis that could lead to development of new drugs effective in considerable portion of the disease cases in a population. Genetic architecture of common diseases is beginning to reveal an incredible diversity of potential genetic causes that act through somewhat limited number of mechanisms with important contribution of environmental interactions. In light of these findings, we present current understanding of genetic architecture of a spectrum of human diseases. We address the encountered problems in susceptibility gene identification, review the success of leading gene identification strategies and discuss current prospects for translating genomic advances into measurable public health benefits.

Key words*: human genomics, public health, disease burden, common complex diseases of late onset, feasibility*

Introduction

Human health is defined by the World Health Organisation as »...a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.«[1]. Critics of this definition argue that health is a process rather than

Received for publication September 10, 2004

a state, and that well being cannot ever be expected to be »complete«[2]. This shows how difficult it is to define either health or disease and to clearly separate them. Human diseases represent a complex spectrum of health disturbance, ranging from those presenting early in life to those with very late onset and from mainly genetically determined to nearly entirely caused by environmental exposures[3].

It is certain, however, that the aetiology of some diseases is considerably simpler than of the others. Some diseases present in first days of life, limiting the importance of environmental exposures (apart from intrauterine) in their aetiology. Such diseases are usually due to structural or functional deficiency of a single protein, resulting from random change in genetic code controlling the synthesis of this protein[4]. The deficit of a single enzyme, or a building protein, eventually leads to cascade of events presenting as disease phenotype. In this case, in all affected patients the disease phenotype is a direct consequence of a single change in genetic information. At the other end of the complexity spectrum are diseases such as cardiovascular disease and cancer. They develop for years as a result of the combination of inherited polygenic susceptibility and lifetime exposure to the environment[3]. Their slow development can only be monitored as a gradual breakdown of physiological mechanisms that became unable to compensate for microscopic structural damage or functional changes at various levels of human organism. The result is a continuing deviation of monitored metabolic or biochemical parameters from the expected value. Over the course of time, this results in the first presenting symptoms and eventually leads to the development of characteristic disease phenotype[3].

The investigation of genetic and environmental factors that interact in causing a disease has been a fundamental goal of modern epidemiology of non-communicable diseases. This science recognises that each human being has different genetic build-up (except monozygotic twins) and personal history of lifetime environmental exposures, and therefore represents unique evolutionary experiment in time and space. However, by designing appropriate prospective and retrospective studies, modern epidemiology attempts to identify genetic characteristics and environmental exposures that are significantly more frequent among the persons with disease phenotype than in general population. Such characteristics, found among those who exhibit disease phenotype more frequently than it would be expected by chance, are then considered »disease risk factors«. Once the risk factors are identified, evaluation of the relative contribution of each of them to the development of disease and assessing whether the associations found between diseases and risk factors are causal or coincidental represent other important goals of modern epidemiology[5–8].

To produce results of any validity, epidemiological studies heavily rely on investigator's ability to precisely measure suspected risk factors (genetic characteristics or environmental exposures) and correctly classify study subject according to the presence or absence of the disease. Presence of disease is typically »measured« by a set of diagnostic criteria. Those criteria have validity of nearly 100% for many diseases (especially if based on pathohistological examinations or various biochemical tests). Still, for some common diseases, misclassification is a serious concern even nowadays[9,10]. Regarding the risk factors, the researchers usually investigated those that could be measured to any degree of precision. This included a multitude of variables related to individual's environmental exposures, physiological or metabolic monitoring, or even psychological profiling. The most investi-

gated risk factors in non-communicable disease epidemiology during the 20th century therefore included person's age, marital status, occupational exposures, smoking, body mass index, blood pressure, cholesterol levels, food intake frequency, psychological assessment of life quality, etc. This all fostered development of environmental, ecological, occupational, nutritional and psychological epidemiology[5].

Genetic epidemiology, however, was unable to expand in a similar fashion. This was because during most of the 20th century it was hardly possible to measure individual's genetic build-up in any direct way. Investigation of genetic risk factors was limited to twin studies[11], heritability studies[12], inbreeding studies[13] and studies of phenotypically expressed genetic variation (e.g. classic erythrocyte antigens, HLA systems)[14,15]. Those studies had a variety of designs and could only provide indirect insights into general patterns of association between factors related to inheritance and occurrence of diseases. This was very far from designs that researchers ideally desired to apply. However, this situation dramatically changed during the past two decades, when first methods for the analysis of the variation in the DNA molecule in each human individual were introduced[16,17]. Such development immediately attracted epidemiologists of non-communicable diseases, who realised the opportunities provided by directly measuring genetic material as a disease risk factor[6]. The early excitement and optimism were justified, as today it is possible to measure individual's genetic build-up more precisely than we will ever be able to measure person's environmental exposures, such as nutrition or psychology. The progress in genomic research, especially in understanding the patterns of variation in the human genome, is destined to nurture further rapid expansion of genetic (genomic) epidemiology, possibly to the levels that eventually may surpass all historic successes of other approaches to epidemiological research in non-communicable diseases[18–21].

In this text, current strategies to find genes responsible for susceptibility to human diseases will be discussed, especially those associated with the greatest burden of disability and death, such as cardiovascular diseases, cancer, diabetes or psychiatric disorders. The ongoing debates on likely genetic architecture of most common human diseases will be addressed, and the suitability of the available tools to find underlying genes to the existing knowledge gathered from both genetic and epidemiological research of human disease aetiology will be analysed. Finally, based on the results of various applied strategies to find common disease susceptibility genes, current prospects for translating genomic advances into measurable public health benefits will be discussed.

Advances in Genome Research Enhanced Development of Genetic Epidemiology

The main goal of genetic epidemiology is to express genetic build-up of an individual as a (predictor) variable, so that it could be statistically correlated to other variables of interest – disease phenotype and interacting environmental risk factors. The criterion variable in this design can be qualitative (presence or absence of disease phenotype) or quantitative (such as levels of blood pressure, serum cholesterol or blood glucose). The advances in study of the human genome during 1990's through publicly funded Human Genome Project (*HGP*) and privately funded project at *Celera Genomics* made important steps towards achieving this goal[22,23]. The sequencing of the human genome showed that all humans are identical in about 99.9% of their genome sequence. More importantly, however, the genes (which are, in broad terms, the segments

of the genome sequence that contain code for protein synthesis) formed only about 3% of that sequence, while the function of the remaining 97% of the genome remained largely unknown.

After those discoveries, it seemed logical to assume that the entire variation in terms of susceptibility to acquiring or avoiding diseases during life span would have to be due entirely to variation found within those 3% of the coding genome sequence, i.e. the genes. Efforts by *HGP* and *Celera* both estimated that there are about 40,000 genes scattered across the human genome[22,23]. The term »gene«, however, has a number of possible definitions. For the purpose of genetic epidemiological studies »gene« can be considered a location in human DNA molecule, usually several thousand nucleotides (i.e. base pairs) long, in which there is interchangeable repeat of shorter segments of coding sequence (»exons«) and longer segments of non-coding sequence (»introns«). During protein synthesis, the coding elements (»exons«) are first transcribed into »messenger RNA« (mRNA) molecule, while the non-coding elements (»introns«) are cut out. Messenger RNA then travels from cell nucleus to ribosomes where it is being read in »triplets« of nucleotides, each encoding the subsequent amino-acid to be built into the structure of a resulting protein. It is important to note that location in DNA that we refer to as a »gene« jointly assumes the coding information inherited from mother (one DNA strain) and father (complementary DNA strain). If those two sequences (»gene alleles«) inherited from both of the parents are identical, then a person will be »homozygous« for that gene, and if they differ the individual is »heterozygous«. If both differing sequences (»alleles«) of a gene in a heterozygous individual are eventually expressed in a phenotype, then they act »codominantly«. If only one of them is phenotypically fully expressed and the other is not, then they are »dominant« and »recessive« alleles, respectively[6,22,23].

The defined terms »homozygosity« and »heterozygosity« refer to characteristic of each gene in an *individual*. Genetic epidemiology is also interested in terms that relate to characteristics of genes in the entire *population*. A gene is »monomorphic« if there is only one (»fixed«) allele for this gene present in the entire population. This means that all the individuals in the population will have to be homozygous for that gene, as they will always inherit the same allele, whoever their parents are. It is thought that about 65% of all human genes are monomorphic. Those probably include many genes with regulatory signalling function that determine that we only should have two kidneys, one liver and five fingers on each hand. Other genes are »polymorphic«, as two or more sequence variants (»alleles«), each with population frequency of at least 1%, can be found in the population at their precise genomic location. The differences between humans at their birth are thought to be mainly due to genetic variants found at polymorphic genomic loci[22,23].

We argued that the goal of genetic epidemiology was to precisely measure genetic build-up of each individual and to quantitatively express it as a variable. This means that genetic epidemiologists are interested only in varying part of human genetic material. If we accept that 97% of non-coding human genome sequence has no apparent function, then this part of the genome should not influence disease susceptibility. In addition to that, some two thirds (27,000) of 40,000 human genes should be monomorphic (»human genetic invariance«), and their role in disease susceptibility therefore entirely passive. This intuitively suggests that only the differences in combinations of alleles that can be found at remaining 13,000 polymorphic human genes deter-

mine genetic variation in susceptibility to human disease.

If this is true, it would reduce the aim of genetic epidemiology to identifying different sequence variants (»alleles«) in polymorphic genes that are significantly more frequently present among the diseased than in healthy population. Subsequent goals would then include understanding how the change in sequence (»mutation«), which introduced this new allele, affects protein synthesis, and by what mechanisms does this lead to disease phenotype. Therefore, finding alleles associated with increased disease risk could, in long term, allow genetic testing of individuals at very early age and undertaking preventive measures from childhood to decrease environmental disease risks. Other apparent benefit of finding these alleles is that insights into change in function of the protein they encode could enhance our understanding of complex mechanisms that cause the disease and provide new targets for development of related drugs.

In an ideal genetic epidemiological longitudinal study, all 13,000 polymorphic human genes would be sequenced in the entire human population. All the existing sequence variants (»alleles«) would be catalogued and their population frequencies determined. Then, all humans would be followed up for disease status during their entire lifetime. After correcting for known environmental risks, gene-environment interactions and gene-gene interactions, it would theoretically be possible to assign relative risks and population-attributable fraction of disease incidence in the population to each allelic variant that exists for each polymorphic gene. This would currently be completely impossible, as large funds and years of work by many research groups would have to be spent at the current state of technology to perform this study even in 10 individuals[22,23]. However, the costs of genotyping (i.e. determining existing alleles at different genome locations in DNA from individual subjects) are rapidly decreasing and the speed at which it can be performed is increasing. This led to recent calls to set the target at an affordable price for whole genome sequencing, i.e. setting a long-term goal to reduce the price of sequencing each individual's genome to about $1,000[24]. This is still far from realistic, but hardly anyone in the field would argue that it might not become possible in the future. Genetic screening and individually tailored drug treatment and prevention strategies could therefore become available to everyone who could afford the sequencing of his/her own genome.

Until then, we will have to design alternative approaches to locate genes underlying human diseases and find allelic variants responsible for functional changes that lead to disturbed health. The main problem is correctly associating disease phenotype with the specific allelic variant of a gene located somewhere in a genome. The advances in human genomics offered first tools. These are based on two fundamental properties of the human genome. The first is the existence of several abundant classes of polymorphic sites scattered across the entire genome with precisely determined locations (»polymorphisms«, »genetic markers«), that can be genotyped using polymerase chain reaction[25–28]. The second is a property of the genome called »linkage disequilibrium«, which suggests that markers that are physically close to each other will tend to be co-inherited, and so will the entire genomic sequence between them[29–32].

Polymorphic Markers and Linkage Disequilibrium Enabled Search for Genes Responsible for Human Diseases

The sequencing of human genome produced evidence of general identity of ge-

nomic sequence among humans in about 99.9%. However, this is not to say that all 6 billion people have nearly identical sequence, and that the entire observed variation is due to the remaining 0.1% of the genetic code. It needs to be understood that virtually any of the 3 billion nucleotides in the haploid genome can be randomly changed by mutation at any generation in any individual[33]. However, as any person has 2 alleles for every gene (apart from those on Y-chromosome), the frequencies of such changes at the time of their first appearance will be 1 in 12 billion alleles, i.e. negligible. We speak of »polymorphic sites« in the human genome only when such newly introduced mutations gradually increase in population frequency over a number of generations, and eventually the frequency of the second allele of a previously monomorphic gene reaches frequency of at least 1% in the entire population[33]. Loci with two or several alleles present in the population, each in frequency greater than 1%, have been found at about 0.1% of human genome sequence. These polymorphic sites can be exactly located by applying specifically tailored restriction enzymes. There are several classes of those »polymorphic sites« in the genome (»genetic markers«, »polymorphysms«), but we will only mention two that have potential to be the most important for finding genes responsible for common human diseases.

The first class of polymorphisms is jointly termed »short sequence repeats« (SSR). At some locations in the genome, short nucleotide sequences tend to repeat several times (e.g. »TATATATA« represents a dinucleotide »TA« which is repeated four times; »GTCGTCGTC« is a trinucleotide »GTC« repeated three times). For some reason, when DNA is duplicated during the cell division, the number of repeating times of these nucleotide sequences can be slightly mistaken (e.g. 6 or 8 repeats are transcribed into the DNA of separating cell instead of 7). An obvious analogy is a tired pupil who has to rewrite a very long sequence of numbers from one paper to another for his homework without making a single mistake. When the numbers that follow each other are diverse (e.g. 274930618…) it is easier to do it correctly, as it is simple to go back to the last transcribed number in case of fall in concentration. However, when the sequence involves repeating the same number several times (e.g. 2866666678…), it is easier to miscount how many times does the number »6« repeat and to make a mistake. If the original number sequence and pupil's homework were passed on to hundreds of other pupils for re-writing, it would eventually result in very few mistakes in the parts of the number sequence which is diverse. However, the number of »6« repeats at the particular line in the paper would probably range anywhere from 3 to 10. This is an analogy on how diversity in SSR arose over the course of human history. The locations in the genome where variable number of short tandem repeats can be found are called »STR«-s (from »short tandem repeats«) or »microsatellites«. The exact location of hundreds of those »markers« has been defined in the human genome, which roughly translates into one per each 1 million of nucleotides (*bp*). They are very informative, as not only are they variable in population at each individual location, but their sequence along the chromosomes (termed »haplotype«) is becoming more unique for each person with introduction of each following microsatellite into a haplotype. With each new polymorphic STR added, the probability of any two persons sharing exactly the same haplotype rapidly diminishes, until it is so small that it allows unique identification of each person based on variation found in the DNA[25,26,34].

Other large and promising class of polymorphisms are »single nucletoide po-

lymorphisms« (SNP). These have been one of the major discoveries of the Human Genome Project. It is thought that there should be several million of these polymorphisms scattered throughout the human genome[33]. We remind that each of 3 billion nucleotides in the human genome can mutate (e.g. from »A« to »C«, »T« or »G«), giving rise to a new »allele« of extremely low frequency in the population (which equals exactly one over doubled number of all humans). However, not many of those mutations will increase in frequency over many generations to reach allele frequency of 1% needed for that precise location in the genome to be formally declared a »single nucleotide polymorphysm« (SNP). It should be noted that each individual SNP marker is incapable to subdivide humans into as many categories as each individual STR marker. This is because each SNP locus can theoretically only have four different »alleles« (»A«, »C«, »T« or »G«), one of which is usually highly predominant. At the same time, each STR marker can have numerous alleles (e.g. 10 alleles with 5 to 14 repeats), many of them with quite similar allele frequencies. This makes a haplotype of any 5 consecutive STR markers far more informative, while any 5 consecutive SNP markers may be shared by a large number of humans. However, the advantage of SNP's over STR's is that they are several orders of magnitude more numerous in the human genome, and that one can be found roughly after every 1 thousand nucleotides, while one STR can approximately be found per each 1 million nucleotides. Thus, SNP's should allow for very fine search for genes on sequence segments that have been suspected as candidate for harbouring disease genes. This is often referred to as »fine-mapping« and it uses the property of genome termed »linkage disequilibrium«, which we will attempt to explain in more detail[35,36].

Human genome is diploid in nature, which means that the large majority of its 3-billion-nucleotide sequence exists in duplicate, one copy inherited from each parent. Those two copies coexist and functionally represent a single human genome. The only two exceptions to this rule are the sequence of Y-chromosome (haploid in nature and always inherited from the father) and mitochondrial DNA (always inherited from the mother). The rest of the human genome is diploid, and it has the property of recombination (»crossing over«): exchanging the corresponding complementary segments of the genome sequence between parents' haploid genome sequences in each person. This happens during meiosis, when the recombined 23-chromosome haploid genome does not contain exclusively single parents' chromosomes, but rather a relatively random mixture of maternal and paternal sequence within each chromosome. This has the obvious consequence of breaking down the haplotypes of specific STR or SNP markers that were originally found along the paternal chromosomes. However, polymorphic alleles on a haplotype that are physically close to each other will have smaller probability of separation through recombination from those that are located far apart. Therefore, some alleles will have a probability of co-inheritance greater that the expected of 50%, as if they were on different chromosomes or recombination could occur randomly between them. For such alleles that remain together on a paternal haplotype (»unrecombined«), we say that they are in »linkage diseqilibrium«[29–32].

An illustrative example to explain the relationships between polymorphic markers, a newly introduced mutation (»disease gene«) and linkage disequilibrium in the population over a course of human history is based on the analogy with the deck of 52 cards. Let us suppose that a disease gene arose somewhere in the geno-

me of a single individual. If we had a deck of 52 cards (»polymorphic markers«) to insert at known positions in the genome of this individual, those cards could help us find the responsible gene several generations later, among all the diseased descendants in the population. Let us suppose that we chose to neatly order all the cards by colour and rank along the genome in which the disease mutation initially developed. Then, a Joker (»disease mutation«) is randomly inserted between Queen and Jack of Hearts. Now, in each mating (»generation«) another deck of cards will be placed next to this original deck, with its cards in completely random order, and three cards at random positions from 1–52 will be exchanged between the two decks (»recombination«). Then, our original deck (»haploid genome«) will be passed on to the next generation, in which Joker will again lead to disease. After repeating this procedure 5–10 times (»generations«), it will still be likely that some pairs, triplets or even quintets of cards (»haplotypes«) that originally followed each other by colour and rank in the first deck still do now. This would mean that they are in »linkage disequilibrium«, and that exchange (»recombination«) did not entirely destroy the original card order (»haplotype«). It would also mean that the Joker can probably still be found between (»linked to«) both Queen and Jack of Hearts, or hopefully at least next to one of them, unless they were among the cards exchanged.

However, if the original deck of cards was multiplying after each exchange, and if Joker (»disease gene«) was invisible by being pulled out at some point, we could still locate where it was originally inserted. This is because both Queen and Jack of Hearts would be unusually frequently present at their original positions in many decks that had Joker, while all the other cards would be entirely random. This would make us reasonably sure it had to be located between those two cards. However, after this procedure is repeated long enough, the original order of cards will eventually completely diminish (»decay of linkage disequilibrium over time«): any card could be found at any position. After many thousand repeats of this procedure and multiplying of the resulting deck, we would not be able to find Queen and Jack of Hearts (»polymorphic markers in linkage disequilibrium with disease gene«) at their original position more frequently than expected by chance.

Study Designs of Genetic Epidemiology: Linkage Analysis versus Genetic Association Studies

Based on this understanding and definition of polymorphic markers and their positions, first genetic epidemiological study designs aiming to find alleles responsible for diseases could begin to emerge. Two types of studies were most obvious choices: one based on collection of families in which disease phenotype was clustering (»linkage analysis«), and the other based on sampling of affected individuals and unaffected controls from the entire population (»genetic association studies«). Both of those approaches were made possible by development of so-called »linkage maps«, i.e. sets of hundreds of STR markers intended for »genome-wide scan«. These »genome-wide scans« aimed to identify at least one marker, but preferably the specific haplotype, that would be found in excessive frequency (with extremely high statistical significance) among affected disease cases in comparison to unaffected controls. The controls would come from either unaffected close relatives (»linkage analysis«) or random unaffected individuals from the population (»genetic association studies«). When a marker was found in excess frequency among affected individuals in comparison to controls, it would then be

presumed that it must be physically »linked« to disease gene. Then, the whole »candidate« region around the marker would be fine-mapped by more dense set of markers to identify the precise position of the responsible gene[37].

Historically, »linkage analysis« has proven to be a very successful and powerful approach if several realistic requirements were fulfilled. Firstly, it was necessary to collect a number of nuclear (or extended) families with two or more disease cases. Secondly, it was essential that the affected persons were correctly classified as »diseased« and unaffected as »healthy«. Thirdly, all disease cases in each family needed to be caused by mutations in only one gene, resulting in apparent pattern of disease inheritance (»Mendelian« diseases: autosomal or sex-linked, with dominant, recessive or codominant mode of inheritance). Fourthly, there should be no genetic heterogeneity in disease aetiology, meaning that all the disease cases in all collected families from the population were caused by mutations in the very same gene with unique position in the human genome[6,37].

If these requirements were met, finding the markers linked to disease gene and a »candidate region« around them was reasonably simple. All of the individuals from all of the families, regardless of disease status, would first be »genotyped« by »genome-wide scan«, i.e. the set of several hundreds of STR markers (»microsatellites«) would be determined in each individual. The average distance between those markers used for such »genome-wide scan« would typically be 5–10 centimorgans (see later). Then, suppose that the investigated disease was dominantly inherited and the father and two children in the first collected family were affected, but mother and the third child were unaffected. It would then be closely monitored which markers were shared between the father and the two affected children, but were absent in mother and the third child. As father transmits a random 50% of his genome to each of his children, this single family would probably already exclude more than 70% of studied markers as potential candidates. Then, the second recruited family would be subsequently added to the analysis, followed by all the others and each time further reducing the possible candidate region of gene position. Eventually, only one marker or a short haplotype would remain non-excluded, pointing to the location of a gene responsible for disease phenotype.

If the studied disease was recessive, especially if it was found in an inbred pedigree (which was usually the case), linkage analysis would be even more powerful. The gene would have to be in the parts of the genome of the affected child that is inherited identical-by-descent from both parents[38]. In children of second cousins, only 1.25% of the genome is expected to be shared identical-by-descent, so if 800 STR markers were used to locate the gene, only about 10 of them would remain possible candidates after the analysis of a single case and his parents. Therefore, it was enough in some instances to find only 3 cases from an inbred isolate human population to be able to narrow down a »candidate region« to a single marker. Many genes for Mendelian diseases were found in isolate and inbred human communities, as it was extremely likely that the rare genetic mutation had to be introduced at only one occasion in history by a single founder[39]. Thus, genetic heterogeneity of the studied disease was not a concern, as was the case when the families were collected from general population.

After the candidate marker linked to disease gene is identified, fine-mapping of the surrounding region is performed with more dense set of markers, involving both STR's and SNP's. The density of selected markers for both genome-wide

scan and subsequent fine-mapping of specific gene is always an important issue in any linkage analysis. As base pairs (nucleotides) measure physical distance between positions in the genome, Morgans measure the probability of recombination occurring between them per meiosis (generation). This makes it more appropriate measure of distance between markers when they are used to locate the unknown gene, whereby 1 million base pairs roughly equals to 1 centimorgan (cM)[40]. The use of linkage set with markers that are 5 cM apart meant that there was only 5% chance that recombination occurred between those two markers during meiosis. The initial genome-wide scans are therefore usually performed with markers spanning 5 or 10 cM, as this represents acceptably low probability that the recombination destroyed connection between the marker and disease gene, given the large number of recruited families that would eventually point to the region of interest. Sometimes, however, especially in inbred communities with large regions of linkage disequilibrium, the eventually determined candidate region spans over the entire haplotype of several markers, i.e. several megabases in distance, and harbours hundreds of genes. Therefore, it typically took years until the causal genetic variant was found by fine mapping and its function understood. However, through this procedure a complete understanding of the aetiology of those diseases was obtained, which would be impossible by any other available means.

Genetic association studies are another obvious strategy for genetic epidemiological research, possibly even simpler in design. They rely on a hope that diseases that will be commonly found in population would have common allelic variants causing them. Those common variants should, ideally, all descend from a single mutational event occurring in one person in human history. In theory, this may be possible if harsh selection was introduced. For example, in time of a catastrophic historic epidemics, a person could have had a disease-causing mutation in his genome located very close to (i.e., »in linkage disequilibrium with«) important and rare HLA variant that helped surviving the epidemics. The selection (»epidemics«) would then cause a massive reduction in the population size (»bottleneck effect«), after which the frequency of favourable HLA variant would be dramatically increased in surviving population, but along with it also the unfavourable disease-causing mutation. The other mechanism through which susceptibility variants for common diseases could have reached high frequencies is »antagonistic pleiotropy«: a variant that was positively selected, because it controlled a fitness-improving trait early in life, may have negative health effects in the post-reproductive period[41]. An apparent example of this »thrifty genotype« hypothesis is type II diabetes[42]. It is very likely that humans were exposed to starvation throughout most of their evolution, thus positively selecting genetic variants for slow and highly efficient food metabolism. However, large environmental changes during the past century and general availability of food in developed countries led to epidemics of obesity later in life, which seems to be the main determinant of type II diabetes.

In genetic association studies, all that is thought required to find the genes underlying late-onset complex diseases would be genotyping many disease cases in the population and many unaffected controls, until some differences in marker allele frequencies begin to show significant differences. Therefore, genetic association studies are practically linkage analyses where the entire population is considered a giant pedigree[7,37]. However, it is still a matter of great debate if this approach is efficient, as it is based on the

number of assumptions that in many cases may prove quite unrealistic[8]. This will be further discussed in the chapter on the leading current approaches to find genes for complex diseases.

Recent Understanding of Genetic Determinants of a Spectrum of Human Diseases

A) Monogenic (Mendelian) Diseases

The common feature of »Mendelian« diseases is that their entire phenotype is caused by a rare mutation in a single gene in the genome. Therefore, the segregation of affected individuals in families follows simple Mendelian predictions[4]. The catalogue of known Mendelian diseases is regularly published and there are currently up to 8,000 listed diseases or syndromes[43]. Last decade witnessed great successes in identifying genetic variants underlying about 1,200 of these diseases[11,44,45]. The key property of Mendelian diseases that made this success possible is that causal genetic mutation is both necessary and sufficient for the development of disease. This ensures good correlation between disease phenotypes and underlying genotypes, given the penetrance of genetic effect is high, which is an important requirement for successful linkage analyses or genetic association studies[44,45].

However, the initial successes in mapping variants underlying Mendelian diseases were soon followed by unexpected insight into the complexity of those »most simple« diseases. Two examples that were most extensively studied are retinitis pigmentosa and thalassaemias. Following the mapping of initial variants responsible for the phenotype of retinitis pigmentosa in large pedigrees, it soon became apparent that there are many different genes, perhaps dozens, scattered throughout the genome, that may lead to the same disease phenotype when mutated. The aetiology of this condition proved even more complex when it was recognised that numerous genetic variants that underlie this condition follow very different modes of transmission: autosomal dominant, autosomal recessive, sex-linked dominant and sex-linked recessive[37,46].

Perhaps the most comprehensive insight into the complexity of phenotype-genotype relationships in monogenic disease was given by Weatherall[45]. His studies on the thalassaemias across the world, arising through positive selection as a condition protective against deadly malaria, but based on extremely different genetic mechanisms, showed how a remarkable diversity in phenotypes is encountered even in this relatively »simple« disease[45]. Thalassaemias are probably the commonest human monogenic diseases, and approximately 7% of world's population are carriers for different inherited disorders of haemoglobin. The extreme phenotypic diversity of this condition encountered throughout the world is determined by »*...layer upon layer of complexity*«[45]. Firstly, there is a variety of primary mutations at the beta-globin genes, similarly to the example of retinitis pigmentosa. Then, there is the action of two known »secondary genetic modifiers«: alpha and gamma-globin genes, which affect the magnitude of excess of alpha chains. The result of the combined action of primary and secondary modifiers is then affected by an unknown number of less well defined »tertiary modifiers« (e.g. vitamin D receptor, oestrogen receptor, genes implicated in collagen synthesis, the locus for hereditary haemocromatosis, UGT glucuronyltransferase, HLA-DR locus, tumour-necrosis factor alpha, intracellular adhesion molecule 1)[45]. Finally, it is recognized that environment, ethnological and cultural factors also strongly affect the disease phenotype, although the underlying mechanisms are less clear[45]. This all shows the complexity underlying even

the »simplest« of genetically determined diseases and it should be taken into account when the studies searching for genetic determinants of more complex diseases are designed, which is not usually the case.

B) »Oligogenic« Diseases

The initial successes in discovering genetic variants underlying monogenic diseases encouraged further progress with the diseases that showed high heritability and were thought to be simpler in aetiology – »oligogenic diseases«. An excellent example is Hirschprung's disease, which is the most common hereditary cause of intestinal obstruction. The pathogenesis of the disease was roughly understood, and the absence of ganglion cells in the specific plexuses of gastrointestinal tract (myenteric and submucosal) were implicated as a cause[47]. This understanding and the early onset of the disease led the scientists to believe that Hirschprung disease (HD) is mainly genetically determined and of relatively simple aetiology, although the clear Mendelian pattern of inheritance could not be established. The efforts based on linkage analysis in pedigrees with multiple affected cases (»multiplex pedigrees«) led to insights which categorised the disease by genetic aetiology and explained both familial and population risk of the disease[48]. There is the more common »short segment« form (S-HD), influenced by the three susceptibility loci at the chromosomes 3, 10 and 19, that explain the complete population incidence of this form of HD. The gene at chromosome 3 is probably *RET*, which seems to have the crucial role in all forms of HD (but not both necessary and sufficient, as is the case with other genes that cause monogenic diseases). Other forms of the disease (»long-segment« and »syndromic« HD) are more rare and genetically more complex, with coding sequence mutations in *RET*, *GDNF*, *EDNRB*, *EDN3* and *SOX10* genes being implicated in various studies[48].

Another disease that seems to show »oligogenic« determination of susceptibility is perhaps also an unexpected one – the leprosy. Although this disease is infectious, the development of the phenotype seems to be strongly genetically determined. A recent paper by Mira et al.[49] showed how the association of the disease with chromosomal region 6q25 was first implicated in a sample of 86 Vietnamese multiplex families using model-free linkage analysis. The association of the candidate region was then repeated in 208 independent simplex Vietnamese families consisting of both parents and one affected child[50]. Fine mapping using single nucleotide polymorphisms implied a regulatory region shared by genes *PARK2* and *PACRG* as responsible for leprosy susceptibility[49]. The authors further replicated this association in a sample of 975 unrelated cases and controls from Brazil, in whom the same variants also showed significant association with leprosy in a candidate gene study[49]. Another chromosomal region (10p13), earlier implicated in an Indian sample of »paucibacillary« disease cases[51], also showed strong association in a »paucibacillary« subset of Vietnamese cases, but not in the »multibacillary« subset. The authors concluded that variants in *PARK2* and *PACRG* are common alleles that confer susceptibility to leprosy »*per se*« globally, and that variants in 10p13 region are also common alleles that determine clinical presentation of disease as »paucibacillary« or »multibacillary«[49,50].

In recent years, more promising evidence is being gathered suggesting that some other diseases may have reasonably simpler architecture of genetic susceptibility than common complex diseases of late onset. There is recently an increased enthusiasm over the identification of variants underlying the susceptibility of as-

thma[52–55], systemic lupus erythematosus[56] and psoriasis[57,58].

C) Complex Polygenic Diseases of Late Onset

Genetic basis of common complex diseases of late onset, responsible for most of the public health burden in wealthy countries of the world, is currently perhaps the greatest focus of interest of the entire biomedical scientific community. This is partly because identification of common genetic risk variants in human populations would enable genetic screening and possibly provide new therapeutic targets for drugs that could be administered in the same manner to large number of persons at increased risk. As both of those prospects would certainly be extremely lucrative and lead to unprecedented increase in revenues for those producing genetic tests and drugs, the investments into search for genetic determinants of common late-onset diseases have been enormous during the recent years. However, the output to date was hardly proportional to the investments. It appears that apparent successes in mapping genes for monogenic diseases and sequencing of the human genome prompted large number of research groups, as well as both private and public investors, into the »gold rush« (search for the »Croesus Code«) that may have been based on slightly overoptimistic assumptions[59]. Primarily, the common diseases of greatest interest, i.e. cardiovascular disease, cancer, type II diabetes and psychiatric disorders, are frequently extremely complex phenotypes that are, contrary to most monogenic diseases, difficult to uniformly measure and define. Secondly, many of the approaches neglected cumulative effects of the environment on disease development, being interested only in genetic component, although most of those diseases show rather low heritability and the majority of cases in general population may be due to environmental exposures. Thirdly, an appealing concept of »common disease/common variant« (CD/CV) gained popularity among the mainstream researchers in the field, based on the assumption that frequent diseases will be determined mainly by genetic variants common to all affected people from different human populations[60].

The outcomes of the research based on previous assumptions were not spectacular. This is not to say that there were no successes. Due to large investments, many small successes have been made and the research field is rapidly expanding. However, the massive undertaking of poorly designed genetic association studies based on possibly false assumptions resulted in a great number of reported associations of common diseases to numerous genes across the genome, but the substantial portion of published reports is likely to be false-positive. Therefore, the issues of repeatability and interpretation of such associations slowly became nearly as important as conducting the studies themselves[61–63]. It is beyond the scope of this paper to list all the encountered genetic associations, but some general conclusions can be drawn. The variants that show repeatable associations with common diseases in more than one population are usually of very small effect and not always common in populations under study. Those variants are therefore unlikely to individually explain substantial proportion of disease burden in studied population. Other variants that were implicated but not repeated may also be causal, but specific of the population under study (e.g., an unusual gene-environment interaction in studied population). Genetic architecture of common diseases is slowly beginning to reveal a large diversity of potential genetic causes, all of them acting through somewhat limited number of mechanisms, with increasin-

gly appreciated contribution of the environmental interactions[3,44].

Several individual efforts, however, increased our understanding of genetic basis of complex polygenic diseases to the extent worthy of specific mention. We will limit the presented examples to cardiovascular diseases and cancer only, as those two complex diseases are jointly responsible for up to 75% of deaths in western countries and therefore represent the principal interest. The first example is the study by Ozaki et al.[64], in which about 1,000 cases of myocardial infarction (MI) were compared to roughly as many control individuals using 92,788 gene-based single-nucleotide polymorphism markers (SNP). The authors used this impressive number of markers in a nearly ideal high-tech genetic association study, conducted in a relatively genetically homogenous Japanese population. Although they covered about half of the entire human genome with their SNP markers, they could only find one statistically significant association (coding region of LTA gene on chromosome 6) when recessive mode of inheritance was assumed, and no significant association ($p<10^{-6}$) under a dominant model. The increase in risk of variant carriers was modest, i.e. about 1.7. Although presented as a success, this study was actually rather discouraging for the proponents of genetic association studies that are based on extremely large numbers of SNP-markers, reasonable samples of cases and controls from the outbred general population and linkage disequilibrium. Even if the reported association is truly causal, it is certain that there must be far more genes underlying MI risk, but they were not identified even in this exercise that was massive in scale.

Another interesting effort is the one performed by deCODE Genomics company in Iceland population. This company was founded in 1996 with aim to identify the genetic causes of common diseases and to develop new drugs and diagnostics based upon its research. It used different approach from the one presented above, based on appreciating the large genetic heterogeneity and complexity underlying common diseases. Iceland population was chosen as it offered most of the potential advantages needed to tackle this complexity – reduced genetic diversity, available disease data and reliable genealogical information. deCODE invested hundreds of millions of dollars into attempts to identify major genes involved in more than 20 of the most common diseases, and has successfully isolated genes in seven of these to date, which is possibly the greatest success rate by any group in the world. The two very recent examples related to cardiovascular disease are identification of the gene encoding phosphodiesterase 4D on chromosome 5 as a risk factor of ischaemic stroke[65], and the gene encoding 5-lipoxigenase activating protein on chromosome 13 as a risk factor for MI and stroke[66].

The investigations of genetic basis of cardiovascular diseases are still at reasonably early stage. However, the research into genetic changes found in human cancers has been conducted for decades[67]. The extreme diversity and complexity of causes, mechanisms and consequences underlying malignant transformation of human cell is possibly a good predictor of what will be encountered in the future when studying genetic basis of other complex diseases[68]. It is now known that familial (monogenic) forms of cancer, such as breast cancer cases »exclusively« due to BRCA1 and BRCA2 mutations, account for only about 20% of the familial breast cancer cases, while familial cases constitute only about 5–10% of all breast cancer cases in general population. Even among those »monogenic« breast cancers, only 25% can be explained by changes in BRCA and other known »breast cancer« genes, while the remaining 75% of famil-

ial cases are due to unknown familial predisposing genes. Non-familial cases, which constitute 90–95% of cases in general population, can therefore be explained only through interaction of unknown polygenic predisposing genes and environmental factors[69].

Some of the changes in genetic material that are frequently postulated as occurring in tumour cells, although neither necessary nor sufficient in all cases in population to lead to cancer, are mutations in coding or regulatory sequences, changes in overall ploidy, high amplification, structural rearrangements and loss of heterozygosity[67]. The key feature of malignant cells is genomic instability, which can be due to inherited mutations in genes that monitor genome integrity, or acquired in any somatic cell during the development of cancer[70]. However, the processes that follow are mediated through and extreme diversity of mechanisms, where it is difficult to distinguish the changes that led to cancer from the changes that arose as a consequence of cellular transformation. The amount of published results in cancer research that is becoming available on different molecular genetic aspects of the disease in recent years is so vast, that possibly the leading current problem seems to be integrating and coordinating this knowledge[68]. It is hoped that many discovered signaling pathways act in parallel through organized networks, but the only way to find those universal principles that are somewhat more limited in number is to combine models and experiments. To achieve this, developing the system of categorization of knowledge will be essential, and one such effort is represented in the National Cancer Institute's Cancer Genome Anatomy project[68]. It is probable that the experience with cancer genetics and genomics will soon be repeated through research into genetic causes of other complex chronic diseases of late onset that we did not mention here (e.g. psychiatric disorders, type II diabetes, and others).

Current Understanding of Genetic Architecture of Common Complex Diseases

There is still a lot of uncertainty and a great deal of controversy over understanding of genetic architecture underlying complex chronic diseases of late onset. These diseases occur mainly in post-reproductive period, and their genetic determinants are therefore less subjected to selective impacts from the environment than is the case with more simple (monogenic and oligogenic) diseases of early onset. However, cumulative negative impacts of the environment over time are also more important in aetiology of late-onset diseases than in early onset diseases. Late-onset diseases are therefore not only genetically more complex, but also multifactorially determined. The key questions that gave rise to recent debates are about the frequency of the responsible susceptibility variants in a population (common / rare), on the number of loci in the genome that underlie these diseases (oligogenic / polygenic), and on the size of their effects (large / small).

Some argue that, because the diseases of late onset are quite common in a population, their genetic determinants (variants responsible for increased susceptibility) should, intuitively, also be common and therefore evolutionary rather old. This is known as the »common disease – common variant« hypothesis (CD/CV)[60]. If this were true, genetic association studies would be expected to be successful and to lead to identification of susceptibility variants. Others, however, argue that, although counter-intuitively, common diseases are more likely to be caused by highly complex interaction of numerous genetic variants, most of them very rare, interacting among themselves and

with the environment. This hypothesis is known as »common disease – rare variant« hypothesis (CD/RV), and would largely undermine currently proposed efforts to identify disease susceptibility genes using genetic association studies[3].

Let us begin with what is known on patterns of human genetic diversity, as it is thought that diversity in variants we carry in our genomes makes some of us susceptible to specific late-onset diseases. Firstly, more than 99% of our genome sequence there is practically no diversity, and the variants at those loci are fixed (i.e. have a population frequency of 100 %). However, the figure of 100% is not entirely accurate, as it is possible that virtually any single nucleotide in this »invariant« part of the genome may be changed (»mutated«) in any individual. However, this is not considered a true polymorphism, as such mutations have incredibly low population frequencies, i.e. practically one in a number equaling twice the total human population size for those occurring the first time in autosomal part of the genome. Such newly arisen single nucleotide polymorphism (SNP) would have to increase its frequency over the course human population history from 1 in hundreds of millions of people to 1 in 100, e.g. 6–7 orders of magnitude, to become a true polymorphism in the population. If that occurred, this SNP can be considered a »common« variant. It is predicted that about 12 million single nucleotides in the genome, i.e. less than 1%, should be polymorphic[33,71]. This magnitude of increase in population frequency for the newly introduced mutations should only be possible through 2 general mechanisms: longtime random genetic drift, or positive selection favouring the carriers in each new generation due to beneficial effects of such mutation in a pre-reproductive life.

Going back to the »monomorphic« (»invariant«) majority of the human genome, it is possible that even this part could confer the susceptibility for development of late-onset diseases. In that case, all the humans would eventually get the disease after being exposed to their environments for sufficiently long periods. The difference in age of onset of the disease cases would be determined solely by cumulative exposure to environmental risk factors during lifetime. This is a »common disease-fixed variant« hypothesis (CD/FV), and there are good examples and plausible explanations why this would be the case for some diseases[42]. For example, it is very likely that starvation was a major selective pressure during most of human history, and that selection strongly favoured new variants that were protecting humans from hunger through more efficient food metabolism. If those variants became fixed, and everyone in the present human population possesses them, the large environmental change in which food became easily accessible in supermarket chains of the western world over the past 50 years would be expected to lead to a pandemic of obesity. This scenario is indeed being observed nowadays. It is likely that atherosclerosis is another example of an »universal« disease, the development of which depends only upon the sum of environmental effects during lifetime.

Two important implications of »CD/FV« hypothesis should be noted. Firstly, it is useless to search for extremely rare variants in this nearly »invariant« part of the genome that would additionally increase risk for e.g. obesity or atherosclerosis above the »universal« risk shared by everyone. This is because population attributable fraction of disease cases due to those specific variants would be negligible and would not lead to any feasible prevention of treatment strategies. Secondly it is apparent that changing behaviour and reducing risky environmental exposures would have much larger public health effects than any improvement in understanding of genetic basis of disea-

ses under CD/FV hypothesis. It is possible that fixed variants indeed do play an important role in genetic architecture of many common complex diseases of late onset, which could partly explain low genetic variance and high environmental variance in many of complex traits and diseases.

However, for other complex diseases of late onset, such as some types of cancer, psychiatric and neurological diseases, it is clear that a significant heritability can be noted, and it is improbable that all the humans would eventually develop those diseases after enough time. In such diseases (e.g., breast cancer, manic depression or multiple sclerosis), genetic factors are likely to play an important role in disease predisposition. As there is variation among humans in their predisposition to developing those diseases, it is thought that this variation is mediated through polymorphic sites in the genome. The key question, however, remains whether the predisposition to disease is a result of action of variants at several loci (oligogenic genetic architecture), all of which carry reasonably large relative risk (e.g. RR>2.0) and are common in a population (CD/CV hypothesis). The alternative hypothesis is that there are many loci across the genome that interact among themselves and with the environment (polygenic genetic architecture), most of which carry very small relative risk associated to individual variants (e.g. RR<1.5) are very rare in a population (CD/RV hypothesis). Under the first model, identification of several responsible variants of large effect would certainly provide clues into disease pathogenesis, and enable genetic screening, prevention and gene-based therapy. Under the second model, the identification of individual rare genetic variants that marginally increase disease risk would contribute very little to understanding of disease pathogenesis and would not lead to feasible diagnostic and therapeutic advances.

The two hypotheses are not necessarily mutually exclusive, and there are arguments to support both. Lohmuller et al. reviewed the replicated gene-disease associations in the world literature, the associated relative risks and the frequencies of the implicated variants in the population, and concluded that there is support for CD/CV hypothesis[72]. However, the associated relative risks were usually overestimated in first published reports and they appear rather small, so that the identified associations largely failed to improve the understanding of disease pathogenesis. It is still thought, however, that the general lack of success in mapping complex disease genes is due to most of the current studies being underpowered (using too few cases and genomic markers to detect associations), and that improved designs and meta-analyses should detect more common variants[8,73,74]. Others doubt that even the increase in number of subjects or number of markers should necessarily help[3]. They argue that if the variants became common in a population, they are likely to either be neutral during pre-reproductive period (and thus increased in frequency by genetic drift), or to have beneficial effects on fitness in prereproductive period (and therefore be positively selected). This would imply that common variants with detrimental effects in post-reproductive period would have had to be evolutionary very old and neutral or even beneficial (»antagonistic pleiotropy« hypothesis) in early life. The authors consider this unlikely, based on summary of the evidence from experimental organisms, or at least not a leading mechanism of common disease pathogenesis. Finally, there is a third scenario in favour of CD/CV hypothesis which cannot be easily dismissed. It hypotheses that some very rare variants became extremely useful in times of large pandem-

ics of infectious diseases, and rapidly increased in frequency over shorter periods of human history. A detrimental variant that was physically close (in tight linkage with) the protective variant could then also increase in frequency via »hitch-hiking« effect, as its detrimental effects on fitness were considerably smaller than the beneficial effects of the linked protective variant under the selective pressure of epidemics. This could explain at least some of the numerous reported associations between specific HLA groups and some relatively common human diseases[75].

The proponents of CD/RV hypothesis use mainly arguments that rely on decades of fundamental research in population genetics and human evolution. As the human population underwent a massive expansion over the past several generations, modelling the predicted number of newly arisen mutations during recent human history implies that the majority of genetic variants contributing to current human genetic and phenotypic variation is predicted to be rare[3]. This argument has recently been strengthened by the discovery that the estimate of the number of mutations per generation per gamete could have been historically underestimated by an order of magnitude[76]. The recent work by Cohen et al. is the first highlighted paper that empirically shows that a risky tail of the distribution of a complex quantitative trait – HDL cholesterol – is determined mainly by rare variants at the population level[77]. Although it may seem counter-intuitive to some scientists (and certainly less attractive for industrial investment) that common diseases of late-onset are mainly caused by a large number of rare variants with small effects, this long neglected view appears to finally be gaining some support.

To summarise the current state in this debate, it is generally accepted that the allelic frequencies in the population and their effect size have an »L«-shaped distribution. The alleles with very large effects, that could provide new insights into disease pathways and mechanisms, are predicted to be very rare in the population. At the same time, the alleles with tiny effects in pre-reproductive period are allowed by selection to become more common[78]. Unless antagonistic pleiotropy is a very general mechanism, or many selectively neutral variants tend to become deletorious in post-reproductive period, it is most likely that ageing process, accompanied with the development of complex chronic diseases of late onset, is mainly mediated by »mutation accumulation« hypothesis. Under such model, the effects of many rare and possibly some common variants interacting among themselves and with the environment would cumulatively lead to the breakdown of intrinsic compensation mechanisms of human organism and eventually manifest as the disease phenotype.

Leading Current Approaches to Identify Common Complex Disease Genes

Based on everything discussed so far, is can be concluded that two main comprehensive approaches to identify common complex disease genes are emerging in the post-genome sequence era. The first approach, which currently has a role of the mainstream research due to large number of publications of its proponents in high-profile journals in recent years, is advocating high-spending efforts in general population of western countries, such as U.K. and U.S.A. (»BioBanks«)[79,80]. Such studies would attempt to generate large population cohorts (of up to 1 million people) with great quantities of information on individual genetic background and environmental exposures. Then, massive genetic association studies would be de-

signed with tens (or even hundreds) of thousands of affected cases and unaffected controls. Their genomes would be sequenced using the large number of single-nucleotide polymorphisms (SNP), which may run into hundreds of thousands of markers per person. The key assumption that would eventually determine the success of this general approach is that the genes underlying most common complex diseases and their underlying haplotypes are common in the population (common disease/common variant hypothesis, »CD/CV«). Recent meta-analysis of genetic association studies published to date was supportive of significant contribution of common variants to common disease susceptibility[72]. The leading current effort following this direction is the »International Hap-Map Project«[81]. This project assumes that most of the human genomic diversity is common, organised into distinct »haplotype blocks« which are also common, and so are the disease susceptibility mutations arising on those blocks. Ultimately, the catalogue of all variants of haplotype blocks in the human genome would enable associating them with common diseases, which would be more feasible strategy than genome-wide scan if the initial assumptions are correct. Moreover, each haplotype would be defined by a minimum informative number of SNP's needed to distinguish it from other haplotype variants (»haplotype tagging«), which would greatly reduce the costs and effort of genotyping[82].

An alternative approach is based on assumption that the key to success in mapping complex disease genes will be through decreasing their aetiologic heterogeneity and improving correlation between genotypes and phenotypes in population under study. This approach advocates the use of isolate human populations with defined number and origin of founders, known ethnic history, possibility to define disease phenotypes and reconstruct individual genealogical records[39,83]. Some of the obvious advantages of this approach is that it is orders of magnitude less costly, and that linkage analyses and genetic association studies can be performed at the same time to support each other and increase the power of the study. However, the main advantage may be that this approach should work even if the variants underlying common complex diseases are rare in general population. This is because such rare variants with large effects may still be common enough in an isolate population to be detected by genetic association study. If they are also rare in an isolate population, they still may be detected by linkage analysis, through an approach that is similar to mapping of monogenic (Mendelian) diseases in isolate populations, which already proved successful in the past. Therefore, even if most of the genetic diversity underlying common diseases proves to be rare, which is somewhat counter-intuitive but predicted by population genetic theories (common disease/rare variant hypothesis, »CD/RV«), the variants could still be identified[3]. The problem with studying isolate populations is that the results may not be relevant for and applicable to wider, general populations. However, it seems that this approach is recently being taken more seriously in the research community due to recent successes in Iceland[65,66]. This is especially true because Iceland's most recent success, finding a gene that increases susceptibility to myocardial infarction, was mapped on a rather common haplotype in Iceland. The association with the suspected genomic location was later confirmed in the population of United Kingdom by candidate gene approach and genetic association study, but with different and less common haplotype involved in the latter population, which casts doubt if the initial finding would be possible in the U.K. population. The results from other

isolate populations that are currently under study are eagerly expected, which include populations of Newfoundland[84], Saami[85], Sardinia[86], Israel[87], Netherlands[88], Croatia[89–96] and Dagestan[97], to mention a few.

In addition to those two leading approaches that are attempting to discover genetic variants underlying complex common diseases in a comprehensive way, and to find genes for many different diseases and traits within the same study, there are also other, more specific approaches that led to important advances. Many disease-oriented groups throughout the world formed multi-centre initiatives to gather large number of patients with a specific diagnosed disease of interest. In such cases, unlike in the two approaches mentioned above, the recruitment of the disease cases is not population-based, but rather hospital-based. When an adequate number of cases is recruited for the study, the gene is initially sought for by either »transmission-disequilibrium test« (TDT) in disease cases only, which requires DNA from both parents regardless of their disease status, or by linkage analysis in expanded pedigrees of the cases. The latter approach is especially powerful if there are multiple cases in the families of the recruited patients, and if the diseases in collected cases were of early onset. When a significant »LOD-score« (statistical measure of association between genomic locus and measured phenotype) is obtained at one or more loci in the genome (the threshold is usually set at LOD=3.0), those regions become »candidate regions« that may harbour a gene responsible for disease susceptibility. The next step is recruiting another large and independent sample of disease cases in the same hospitals and checking if those »candidate regions« are again associated with the disease status in this new sample of patients, thus reaffirming or dismissing the importance of the loci implicated in the first study. This general approach, although only single disease specific, subject to number of confounding effects (such as population stratification)[98]) and often disregarding environmental effects, has still been quite successful. Many of the repeatable associations in the current literature were reported after this initial approach[99,100].

Translating Genomic Knowledge into Public Health Benefits

There are two main expectations from genetic epidemiology in terms of delivering results that would have major impact on public health. The first one is the association of different genetic variants with specific health risks and the translation of that knowledge into development of commercially available genetic tests that could predict diseases. The second one is the understanding of disease mechanisms and obtaining new insights in disease pathogenesis, which would reveal new targets for development of drugs that could prevent or reverse the course of human diseases[101]. Although several years ago those targets appeared far from reality, today there is a growing economic sector of biotechnology, in which large number of private companies are attempting to deliver one or both of those goals, and some tests and genome-based drugs are already being offered on the market. We will briefly address the current status of advancement towards those two main goals of genetic epidemiology.

In terms of genetic testing, recently also called »genomic profiling«, it is based on an expectation that knowing most of genetic variants that could increase the risk of disease would enable the development of »DNA chips« containing this information. Those chips could then scan for the presence of extremely large number of such variants in any individual's genome at birth. After the scan of the ge-

nome, the chip would compute the lifetime risk of various diseases, thus being a powerful tool of a »personalised medicine«. Although a number of private companies already offer genomic profiling for »oxidative stress«, »susceptibility to obesity or osteopenia«, »nicotine or alcohol dependance«, etc., these could hardly have any scientific basis, as the genetic architecture of those traits and responsible variants are simply not known with any accuracy at present. Therefore, the problem of regulating the marketing of such tests is growing recently, as it is entirely unlikely that we could have useful and reliable genetic tests that could predict individual risk of common complex diseases in foreseeable future[101]. However, in the meanwhile it is certain that the ease of marketing of those tests (internet) coupled with the desire of consumers in some western countries to actively control their health at any cost may result in creation of smaller market for these tests of unproven value. Similar or even more dramatic examples have already been seen with the popularity of various diets and food supplements[20,21,102,103].

However, although predicting complex diseases in individuals based on their unique genome sequence may still be far from reality, there have been some positive developments in achieving the second target – associating genes with diseases to understand aetiology which eventually led to new drug discoveries. The two frequently cited examples are imatinib and trastuzumab[104,105]. Imatinib (Gleevec) was developed following the discovery that a chromosome translocation created a new gene structure in some patients with chronic myeloid leukaemia, and the drug binds to the protein product of this gene and fights disease progression where other treatments fail[104]. Trastuzumab (Herceptin) did not appear to significantly improve survival of breast cancer patients, until it was realised that it is very efficient, but only in a subset of Her-2 positive breast cancer cases. Although those examples based on molecular understanding of disease pathogenesis may seem spectacular, the more general view is that the successes in finding the genes underlying common complex diseases have been very modest in relation to unprecedented investments into this research from both industry and academic community. Although numerous associations of various genes with a spectrum of diseases have been reported, only a few of those associations have been replicable. Moreover, the risks associated to implicating variants were usually very modest and promising little hope for contributing to improvement in understanding of disease aetiology. As the current level of investment in unsustainable, it is becoming apparent that the successes in mapping genes for Mendelian diseases will not be easily repeated with complex diseases, and that more rational strategies for associating genes and disease phenotypes will need to be developed.

Based on this premises, the recent review by Merikangas and Risch[106] provided a more sober assessment of the current status of search for common complex disease genes and strategies for future investments. The authors argue that investments are justified for the diseases that are: (i) common in the population (associated with substantial public health burden); (ii) can be precisely diagnosed (to avoid misclassification of cases, which dramatically reduces power of genetic association and linkage studies); (iii) show substantially increased risk in relatives of diseased cases (to demonstrate the role of genetic effects as opposed to environment); and (iv) for which no preventable environmental risks are known. It is worrying that enormous funds are being invested in searching for genetic basis of diseases or conditions that hardly show any heritability, cannot be diagnosed with

any precision, or for which the funds would be far better placed in fighting environmental risks rather than searching for genetic clues. For example, investing funds in finding genes for increased individual »nicotine dependence« or »alcohol dependence« is entirely misplaced, as those traits have been shown to cluster more strongly in social groups of different genetic background rather than in families. In addition, the benefit of public health intervention on reducing nicotine and alcohol consumption in a population, which are cheaper than gene searches, far outweigh any possible benefit that could come out of knowing genes that predispose an individual to alcohol consumption or smoking. The large majority of cases of coronary heart disease or diabetes type II in population can be explained by environmental risks such as unhealthy diet, lack of physical activity and smoking. It is unlikely that finding genetic variants that mildly increase risk of those diseases, and even developing therapies based on that knowledge, would lead to appreciable decrease of their public health burden in a population.

The examples of diseases where genomic revolution could prove helpful, however, are Alzheimer's disease, multiple sclerosis, autism or schizophrenia, where relatives of diseased are clearly at greater risk, there are no known preventable environmental risks and which are sufficiently common in the population to justify large investments. For other diseases, the funds would be better placed into research of determinants of human behaviour and motivation for leading more healthy lifestyles. This is all particularly relevant for the population of western countries, where an estimated 150 million people already have type II diabetes and are overweight. However, only a minority of world's population lives in developed countries. In recent years, calls have been made upon international scientific community not to forget the majority of world's population that does not represent a lucrative market for pharmaceutical industry, but could also perhaps benefit from new genomic and molecular technologies, even more than the western world[107]. The facts that 11 million children under five years still die annually of mainly preventable or easily treatable causes, such as pneumonia, diarrhoea, malaria or malnutrition, and that more than two thirds of people with AIDS live in countries with virtually non-existent health systems, are more than worrying. Those people could greatly benefit from recombinant vaccines that use genomic technology, or also from molecular tests that could precisely diagnose the aetiology of their infections and thus enable more efficient use of sparse medicines available to those populations. It remains to be seen whether the genomic revolution of 21st century will truly revolutionise medicine and result in major public health benefits for all of the humanity. The alternative scenario is that it may only deliver partial successes which will become available to the rich minority and thus further increase the gap between the world's rich and the poor, as was the case with recent revolutions in informatics and telecommunication technologies in 1980's and 1990's.

Acknowledgements

This paper was partly supported from the research grants of the Croatian Ministry of Science, Education and Sport to I.R. (0108330) and P.R. (0196005), The Wellcome Trust (IRDA) to Professor Harry Campbell (H.C.) and I.R., and The Royal Society UK to H.C. and I.R. I.R. gratefully acknowledges support from the University of Edinburgh, Medical Research Council UK and Overseas Research Scheme UK.

REFERENCES

1. WORLD HEALTH ORGANIZATION: Preamble to the Constitution of the World Health Organization. (Official Records of the World Health Organization No. 2, Geneva, 1948). — 2. BACHMANN, W. Munch. Med. Wochenschr., 119 (1977) 349. — 3. WRIGHT, A., B. CHARLESWORTH, I. RUDAN, A. CAROTHERS, H. CAMPBELL, Trends Genet., 19 (2003) 97. — 4. HAMOSH, A., A. F. SCOTT, J. AMBERGER, D. VALLE, V. A. McKUSICK, Hum. Mutat., 15 (2000) 57. — 5. ROTHMAN, K. J., S. GREENLAND: Modern epidemiology, 2nd edition. (Lippincott, Williams & Wilkins Publishers, New York, 1998). — 6. KHOURY, M. J., B. H. COHEN, T. H. BEATY: Fundamentals of genetic epidemiology, 1st edition. (Oxford University Press, Oxford, 1993). — 7. CARDON, L. R., J. I. BELL, Nat. Rev. Genet., 2 (2001) 91. — 8. ZONDERVAN, K. T., L. R. CARDON, Nat. Rev. Genet., 5 (2004) 89. — 9. FREIMER, N., C. SABATTI, Nat. Genet., 34 (2003) 15. — 10. MERIKANGAS, K. R., N. RISCH, Science, 302 (2003) 599. — 11. PELTONEN, L., A. PALOTIE, K. LANGE, Nat. Rev. Genet., 1 (2000) 182. — 12. BARTON, N. H., P. D. KEIGHTLEY, Nat. Rev. Genet., 3 (2002) 11. — 13. BITTLES, A. H., J. V. NEEL, Nat. Genet., 8 (1994) 117. — 14. CROW, J. F., Nat. Rev. Genet., 1 (2000) 40. — 15. PASSARGE, E. Color Atlas of Genetics. 2nd ed. (Georg Thieme Verlag, Stuttgart, 2001). — 16. ERLICH, H. A., D. GELFAND, J. J. SNINSKY, Science, 252 (1991) 1643. — 17. HOUSMAN, D., N. Engl. J. Med., 332 (1995) 318. — 18. REICH, D. E., E. S. LANDER, Trends Genet., 17 (2001) 502. — 19. RISCH, N., K. MERIKANGAS, Science, 273 (1996) 1516. — 20. BENTLEY, D. R., Nature, 429 (2004) 440. — 21. BELL, J., Nature, 429 (2004) 453. — 22. LANDER, E. S., et al., Nature, 409 (2001) 860. — 23. VENTER, J. C., et al., Science,.291 (2001) 1304. — 24. SUBRAMANIAN, G., M. D. ADAMS, J. C. VENTER, S. BRODER, JAMA, 286 (2001) 2296. — 25. RUBINSTEIN, D. C., et al., Nat. Genet., 10 (1995) 337. — 26. COLLINS, F. S., M. S. GUYER, A. CHAKRAVARTY, Science, 282 (1998) 682. — 27. BROWN, T. A.: Genomes. (Bios Scientific Publications, Oxford, 1999). — 28. SACHIDANANDAM, R., et al., Nature, 409 (2001) 928. — 29. REICH, D. E., M. CARGILL, S. BOLK, J. IRELAND, P. C. SABETI, D. J. RICHTER, T. LAVERY, R. KOUYOUMJIAN, S. F. FARHADIAN, R. WARD, E. S. LANDER, Nature, 411 (2001) 199. — 30. ARDLIE, K. G., L. KRUGLYAK, M. SEIELSTAD, Nat. Rev. Genet., 3 (2002) 299. — 31. WALL, J. D., J. K. PRITCHARD, Nat. Rev. Genet., 4 (2003) 587. — 32. GABRIEL, S. B., S. F. SCHAFFNER, H. NGUYEN, J. M. MOORE, J. ROY, B. BLUMENSTIEL, J. HIGGINS, M. DEFELICE, A. LOCHNER, M. FAGGART, S. N. LIU-CORDERO, C. ROTIMI, A. ADEYEMO, R. COOPER, R. WARD, E. S. LANDER, M. J. DALY, D. ALTSHULER, Science, 296 (2002) 2225. — 33. KRUGLYAK, L., D. A. NICKERSON, Nat. Genet., 27 (2001) 234. — 34. STRACHAN, T. A., A. P. READ: Human Molecular Genetics, 2nd ed. (Bios Scientific Publications, Oxford, 1999). — 35. MARTH, G., R. YEH, M. MINTON, R. DONALDSON, Q. LI, S. DUAN, R. DAVENPORT, R. D. MILLER, P. Y. KWOK, Nat. Genet., 27 (2001) 371. — 36. CARLSON, C. S., M. A. EBERLE, M. J. RIEDER, J. D. SMITH, L. KRUGLYAK, D. A. NICKERSON, Nat. Genet., 33 (2003) 518. — 37. TERWILLIGER, J. D., H. H. H. GORING, Hum. Biol., 72 (2000) 63. — 38. LANDER, E. S., D. BOTSTEIN, Science, 236 (1987) 1567. — 39. PELTONEN, L., A. PALOTIE, K. LANGE, Nat. Rev. Genet., 1 (2000) 182. — 40. MCVEAN, G. A., S. R. MYERS, S. HUNT, P. DELOUKAS, D. R. BENTLEY, P. DONNELLY, Science, 304 (2004) 581. — 41. CHARLESWORTH, B., K. A. HUGHES, Proc. Natl. Acad. Sci. USA, 93 (1996) 6140. — 42. NEEL, J. V., Am. J. Hum. Genet., 14 (1962) 353. — 43. HAMOSH, A., A. F. SCOTT, J. AMBERGER, C. BOCCHINI, D. VALLE, V. A. MCKUSICK, Nucleic Acids Res., 30 (2002) 52. — 44. BOTSTEIN, D., N. RISCH, Nat. Genet., 33 Suppl (2003) 228. — 45. WEATHERALL, D. J., Nat. Rev. Genet., 2 (2001) 245. — 46. HECKENLIVELY, J. R., S. P. DAIGER, In: RIMOIN, D. L., J. M. CONNOR, R. E. PYERITZ (Eds): Emory and Rimoin's Principles and Practice of Medical Genetics, 3rd Edition. (Churchill-Livingstone, Edinburgh, 1996). — 47. HOLSCHNEIDER, A. M.: Hirschprung's disease. (Thieme-Stratton, New York, 1982). — 48. GABRIEL, S. B., R. SALOMON, A. PELET, M. ANGRIST, J. AMIEL, M. FORNAGE, T. ATTIE-BITACH, J. M. OLSON, R. HOFSTRA, C. BUYS, J. STEFFANN, A. MUNNICH, S. LYONNET, A. CHAKRAVARTI, Nat. Genet., 31 (2002) 89. — 49. MIRA, M. T., A. ALCAIS, V. T. NGUYEN, M. O. MORAES, C. DI FLUMERI, H. T. VU, C. P. MAI, T. H. NGUYEN, N. B. NGUYEN, X. K. PHAM, E. N. SARNO, A. ALTER, A. MONTPETIT, M. E. MORAES, J. R. MORAES, C. DORE, C. J. GALLANT, P. LEPAGE, A. VERNER, E. VAN DE VOSSE, T. J. HUDSON, L. ABEL, E. SCHURR, Nature, 427 (2004) 636. — 50. MIRA, M. T., A. ALCAIS, N. VAN THUC, V. H. THAI, N. T. HUONG, N. N. BA, A. VERNER, T. J. HUDSON, L. ABEL, E. SCHURR, Nat. Genet., 33 (2003) 412. — 51. SIDDIQUI, M. R., et al., Nat. Genet., 27 (2001) 439. — 52. ZHANG, Y., N. I. LEAVES, G. G. ANDERSON, C. P. PONTING, J. BROXHOLME, R. HOLT, et al., Nat. Genet., 34 (2003) 181. — 53. ALLEN, M., A. HEINZMANN, E. NOGUCHI, G. ABECASIS, J. BROXHOLME, C. P. PONTING, Nat. Genet., 35 (2003) 258. — 54. LAITINEN, T., A. POLVI, P. RYDMAN, J. VENDELIN, V. PULKKINEN, P. SALMIKANGAS, et al., Science, 304 (2004) 300. — 55. WILLS-KARP, M., S. L. EWART, Nat. Rev. Genet., 5 (2004) 376. — 56. PROKUNINA, L., C. SASTILLEJO-LOPEZ, F. OBERG, I. GUNNARSSON, L. BERG, V. MAGNUSSON, et al., Nat. Genet., 32 (2002) 666. — 57. HELMS, C., L. CAO, J. G. KRUEGER, E. M. WIJSMAN, F. CHAMIAN, D. GORDON, M. HEFFERNAN, J. A. WRIGHT DAW, J. ROBARGE, J. OTT, P.-Y. KWOK, A. MENTER, A. M. BOWCOCK, Nat. Genet., 35 (2003) 349. — 58. INTERNATIONAL PSORIASIS GENETICS CONSORTIUM, Am. J.

Hum. Genet., 73 (2003) 430. — 59. WRIGHT, A. F., N. D. HASTIE, Genome Biol., 2 (2001) 2007.1–8. — 60. LANDER, E. S., Science, 274 (1996) 536. — 61. CAMPBELL, H., I. RUDAN, The Pharmacogenomics J., 2 (2002) 349. — 62. IOANNIDIS, J. P. A., E. E. NTZANI, T. A. TRIKALINOS, D. G. CONTOPOULOS-IOANNIDIS, Nat. Genet., 29 (2001) 306. — 63. DAHLMAN, I., I. A. EAVES, R. KOSOY, V. A. MORRISON, J. HEWARD, S. C. L. GOUGH, et al., Nat. Genet., 30 (2002) 149. — 64. OZAKI, K., K. INOUE, H. SATO, A. IIDA, Y. OHNISHI, A. SEKINE, H. SATO, K. ODASHIRO, M. NOBUYOSHI, M. HORI, Y. NAKAMURA, T. TANAKA, Nature, 429 (2004) 72. — 65. GRETARSDOTTIR, S., G. THORLEIFSSON, S. T. REYNISDOTTIR, A. MANOLESCU, S. JONSDOTTIR, T. JONSDOTTIR, T. GUDMUNDSDOTTIR, S. M. BJARNADOTTIR, O. B. EINARSSON, H. M. GUDJONSDOTTIR, M. HAWKINS, et al., Nat. Genet., 35 (2003) 131. — 66. HELGADOTTIR, A., A. MANOLESCU, G. THORLEIFSSON, S. GRETARSDOTTIR, H. JONSDOTTIR, U. THORSTEINSDOTTIR, N. J. SAMANI, G. GUDMUNDSSON, S. F. GRANT, et al., Nat. Genet., 36 (2004) 233. — 67. BALMAIN, A., J. GRAY, B. PONDER, Nat. Genet., 33 Suppl (2003) 238. — 68. EDITORIAL, Nat. Genet., 36 (2004) 313. — 69. PHAROAH, P. D. P., A. ANTONIUO, M. BOBROW, R. L. ZIMMERN, D. F. EASTON, B. A. J. PONDER, Nat. Genet., 31 (2002) 33. — 70. LENGAUER, C., K. W. KINZLER, B. VOGELSTEIN, Nature, 396 (1998) 643. — 71. REICH, D. E., S. F. SCHAFFNER, M. J. DALY, G. MCVEAN, J. C. MULLIKIN, J. M. HIGGINS, D. J. RICHTER, E. S. LANDER, D. ALTSHULER, Nat. Genet., 32 (2002) 135. — 72. LOHMUELLER, K. E., C. L. PEARCE, M. PIKE, E. S. LANDER, J. N. HIRSCHHORN, Nat. Genet., 33 (2003) 177. — 73. FREIMER, N., C. SABATTI, Nat. Genet., 10 (2004) 1045. — 74. CARLSON, C. S., M. A. EBERLE, L. KRUGLYAK, D. A. NICKERSON, Nature, 429 (2004) 446. — 75. LARSEN, C. E., C. A. ALPER, Curr. Opin. Immunol., 16 (2004) 660. — 76. DENVER, D. R., K. MORRIS, M. LYNCH, W. K. THOMAS, Nature, 430 (2004) 679. — 77. COHEN, J. C., R. S. KISS, A. PERTSEMLIDIS, Y. L. MARCEL, R. MCPHERSON, H. H. HOBBS, Science, 305 (2004) 869. — 78. PRITCHARD, J. K., N. J. COX, Hum. Mol. Genet., 11 (2002) 2417. — 79. WRIGHT, A. F., A. D. CAROTHERS, H. CAMPBELL, Pharmacogenomics J, 2 (2002) 75. — 80. COLLINS, F. S., Nature, 429 (2004) 475. — 81. THE INTERNATIONAL HAPMAP CONSORTIUM, Nature, 426 (2003) 789. — 82. JOHNSON, G. C., L. ESPOSITO, B. J. BARRATT, A. N. SMITH, J. HEWARD, G. DI GENOVA, H. UEDA, H. J. CORDELL, I. A. EAVES, F. DUDBRIDGE, R. C. TWELLS, F. PAYNE, W. HUGHES, S. NUTLAND, H. STEVENS, P. CARR, E. TUOMILEHTO-WOLF, J. TUOMILEHTO, S. C. GOUGH, D. G. CLAYTON, J. A. TODD, Nat. Genet., 29 (2001) 233. — 83. WRIGHT, A. F., A. D. CAROTHERS, M. PIRASTU, Nat. Genet., 23 (1999) 397. — 84. RAHMAN, P., A. JONES, J. CURTIS, S. BARTLETT, L. PEDDLE, B. A. FERNANDEZ, N. B. FREIMER, Hum. Mol. Genet., 13 (2004) 1287. — 85. KAESSMANN, H., S. ZOLLNER, A. C. GUSTAFSSON, V. WIEBE, M. LAAN, J. LUNDEBERG, M. UHLEN, S. PAABO, Am. J. Hum. Genet., 70 (2002) 673. — 86. TENESA, A., A. F. WRIGHT, S. A. KNOTT, A. D. CAROTHERS, C. HAYWARD, A. ANGIUS, I. PERSICO, G. MAESTRALE, N. D. HASTIE, M. PIRASTU, P. M. VISSCHER, Hum. Mol. Genet., 13 (2004) 25. — 87. SHIFMAN, S., J. KUYPERS, M. KOKORIS, B. YAKIR, A. DARVASI, Hum. Mol. Genet., 12 (2003) 771. — 88. AULCHENKO, Y. S., P. HEUTINK, I. MACKAY, A. M. BERTOLI-AVELLA, J. PULLEN, N. VAESSEN, T. A. RADEMAKER, L. A. SANDKUIJL, L. CARDON, B. OOSTRA, C. M. VAN DUIJN, Eur. J. Hum. Genet., 12 (2004) 527. — 89. RUDAN, I., H. CAMPBELL, P. RUDAN, Coll. Antropol., 23 (1999) 531. — 90. FORENBAHER, S., Coll. Antropol., 26 (2002) 361. — 91. MALNAR, A., Coll. Antropol., 26 (2002) 411. — 92. RUDAN, I., D. RUDAN, H. CAMPBELL, Z. BILOGLAV, R. UREK, M. PADOVAN, L. SIBBETT, B. JANIĆIJEVIĆ, N. SMOLEJ-NARANČIĆ, P. RUDAN, Coll. Antropol., 26 (2002) 421. — 93. RUDAN, I., M. PADOVAN, D. RUDAN, H. CAMPBELL, Z. BILOGLAV, B. JANIĆIJEVIĆ, N. SMOLEJ-NARANČIĆ, P. RUDAN, Coll. Antropol., 26 (2002) 421. — 94. ŠKREBLIN, L., L. ŠIMIČIĆ, A. SUJOLDŽIĆ, Coll. Antropol., 26 (2002) 333. — 95. ŠKARIĆ-JURIĆ, T., Coll. Antropol., 27 (2003) 229. — 96. ŠKARIĆ-JURIĆ, T., E. GINSBURG, E. KOBYLIANSKY, I. MALKIN, N. SMOLEJ-NARANČIĆ, P. RUDAN, Coll. Antropol., 27 (2003) 135. — 97. BULAEVA, K. B., T. A. PAVLOVA, R. M. KURBANOV, S. LEAL, O. A. BULAEV, Genetika, 39 (2003) 413. — 98. FREEDMAN, M. L., D. REICH, K. L. PENNEY, G. J. MCDONALD, A. A. MIGNAULT, N. PATTERSON, S. B. GABRIEL, E. J. TOPOL, J. W. SMOLLER, C. N. PATO, M. T. PATO, T. L. PETRYSHEN, L. N. KOLONEL, E. S. LANDER, P. SKLAR, B. HENDERSON, J. N. HIRSCHHORN, D. ALTSHULER, Nat. Genet., 36 (2004) 388. — 99. LAZARUS, R., B. A. RABY, C. LANGE, E. K. SILVERMAN, D. J. KWIATKOWSKI, D. VERCELLI, W. J. KLIMECKI, F. D. MARTINEZ, S. T. WEISS, Am. J. Respir. Crit. Care Med., 170 (2004) 594. — 100. BONIFATI, V., P. RIZZU, J. VAN BAREN, O. SCHAAP, G. J. BREEDVELD, E. KRIEGER, D. C. J. DEKKER, F. SQUITIERI, P. IBANEZ, M. JOOSSE, J. W. VAN DONGEN, N. VANACORE, J. C. VAN SWIETEN, A. BRICE, G. MECO, C. M. VAN DUIJN, B. A. OOSTRA, P. HEUTINK, Science, 299 (2003) 256. — 101. HAGA, S. B., M. J. KHOURY, W. BURKE, Nat. Genet., 34 (2003) 347. — 102. EGGER, G., G. LIANG, A. APARICIO, P. A. JONES, Nature, 429 (2004) 457. — 103. EVANS, W. E., M. V. RELLING, Nature, 429 (2004) 464. — 104. DRUKER, B. J., Semin. Hematol., 40 (2003) 50. — 105. FORNIER, M., M. RISIO, C. VAN POZNAK, A. SIEDMAN, Oncology, 16 (2002) 1340. — 106. MERIKANGAS, K. R., N. RISCH, Science, 302 (2003) 599. — 107. WEATHERALL, D. J., Science, 302 (2003) 597.

I. Rudan

Department of Medical Statistics, Epidemiology and Medical Informatics, School of Public Health »Andrija Štampar«, School of Medicine, University of Zagreb, Rockefellerova 4, 10000 Zagreb, Croatia
e-mail: irudan@hotmail.com

OD NAPRETKA U ISTRAŽIVANJU GENOMA DO JAVNO-ZDRAVSTVENE DOBROBITI: NEPODNOŠLJIVA LAKOĆA STAJANJA U MJESTU

S A Ž E T A K

Genetske odrednice čestih ljudskih bolesti još uvijek su slabo razjašnjene. Zahvaljujući golemim ulaganjima, učinjen je niz manjih napredaka i znanstveno područje se ubrzano razvija. Međutim, identificirane genetske varijante koje pokazuju ponovljivu povezanost s kompleksnim bolestima su obično slabog učinka. Stoga nije vjerojatno da će pojedinačne genske varijante razjasniti značajan dio morbiditeta u zajednici ili pak pružiti nove uvide u patogenezu bolesti koji bi mogla dovesti do razvoja novih lijekova primjenjivih na značajniji udio oboljelih u populaciji. Genetska arhitektura čestih bolesti počinje razotkrivati veliku raznovrsnost potencijalnih genetskih uzroka koji djeluju kroz ponešto ograničen broj mehanizama uz važan utjecaj interakcija s okolišem. U svjetlu spomenutih otkrića, u ovom smo radu prikazali sadašnje razumijevanje genetske arhitekture cijelog spektra ljudskih bolesti. Osvrnuli smo se na probleme koji su proizašli iz pokušaja pronalaženja gena odgovornih za sklonost bolestima, izvršili kratak pregled uspješnosti vodećih strategija za identificiranje gena, te razmotrili perspektive za prevođenje napretka u genomskim istraživanjima u mjerljivu javno-zdravstvenu dobrobit.