ISSN 1022-7954, Russian Journal of Genetics, 2011, Vol. 47, No. 12, pp. 1395–1405. © Pleiades Publishing, Inc., 2011.
Original Russian Text © V.P. Puzyrev, A.N. Kucher, 2011, published in Genetika, 2011, Vol. 47, No.12, pp. 1573–1585.

THEORETICAL ARTICLES AND REVIEWS

Evolutionary Ontogenetic Aspects of Pathogenetics of Chronic Human Diseases

V. P. Puzyrev[a, b] and A. N. Kucher[a]

[a] *Research Institute of Medical Genetics, Siberian Branch, Russian Academy of Medical Sciences, Tomsk, 634050 Russia*
valery.puzyrev@medgenetics.ru
[b] *Department of Medical Genetics, Siberian State Medical University, Tomsk, 634050 Russia*

Received April 14, 2011

Abstract—This article is a review of scientific publications, in which issues of pathogenetics of multifactorial diseases (MFDs) are considered from the viewpoint of evolution and ontogeny. Concepts explaining significance of evolutionary processes in the formation of genetic architecture of human chronic diseases ("thrifty" genomes and phenotypes, "drifty genes," decanalization) are analyzed. The roles of natural selection and genetic drift in the formation of hereditary diversity of genes for susceptibility to MFDs are considered. The modern concept of "disease ontogeny" (somatic mosaicism, loss of heterozygosity, paradominant inheritance, epigenetic variability) is discussed. It is demonstrated that the evolutionary and ontogenetic approaches to analysis of genimuc and other "-omic" data are essential for understanding the biology of diseases.

DOI: 10.1134/S102279541112012X

INTRODUCTION

Recently, the experts of National Human Genome Research Institute (United States) identified five stages on the course from genomics to using its advances in practical medicine ("from basepairs to bedside"), which are consecutive and overlap [1]: cataloguing the human genome structure (stage 1, 1990–2003), studying the genome biology (stage 2, 2004–2010) as the basis for understanding the biology of the disease (stage 3, 2011–2020), and then the establishemnt of the medical science (stage 4, after 2020) that will considerably improve the healthcare effectiveness (stage 5). According to the experts, one of the important directions for understanding the genome and disease biology is the evolutionary context in considering the advances of genomics, proteomics, and other "-omics."

The importance of the evolutionary approach to the issues of disease and health has been under active discussion [2, 3]. The contribution of Russian researchers (A.N. Severtsov, I.I. Schmalhausen, N.V. Timofeeff-Ressovsky) in the development of the evolution theory was highly appraised. This contribution was again considered in connection with the appearance of a new field of study, evolutionary developmental biology, abbreviated as EvoDevo [4].

The significance of the evolution theory for medicine was studied by Russian geneticists of the next generation, including Yu.P. Altukhov. Altukhov substantiated the concept of the adaptive optimum of genetic diversity as well as the importance of viewing an organism as "a developing hierarchical system of populations of cells and genes interacting in a nonrandom manner" and formulated the view on "the unity of the evolutionary and developmental processes" [5, 6]. These aspects, in our opinion, are important for understanding human genome biology and the formation of genomes susceptible or resistant to diseases. The present review deals with the results of investigations in these directions.

EVOLUTIONARY MODELS OF GENETIC ARCHITECTURE OF MULTIFACTORIAL DISEASES

Genetic architecture of multifactorial diseases (MFDs) is the number of genetic polymorphisms affecting the risk of MFDs, the distribution of their allele frequencies and the intensity of their effects, as well as the genetic mode of their expression (additivity, dominance and (or) epistasis, pleiotropy) [7]. Evolutionary genetic analysis of population dynamics factors (mutation, genetic drift, natural selection), which that potentially can affect the genetic structure of susceptibility (genetic architecture) of MFDs, gives a better idea of the number and frequency of susceptibility alleles, helping to establish the optimal strategy of identification of these alleles. The dependence of analytical approaches on the distribution of genetic variants of MFD susceptibility genes is illustrated by the data presented in Table 1. Note that strategies of identifying susceptibility genes depend not only on the proportion of common and rare variants in the population, but also on localization of these genes in few (oligogenes) or many (polygenes) loci.

Table 1. Dependence of Analytical Approaches from the Incidence of Genetic Variants of MFD Genes

Class type	Frequency of minor allele	Use for analysis
Very frequent	5–50%	Associative analysis using current association techniques for genome-wide analysis
Not frequent	1–5%	Associative analysis using variants listed in the 1000 *Genomes* project
Rare (but not particular)	Less than 1%	Resequencing upon extreme phenotypes as well as analysis of cosegregation in families
Particular	Restricted to probands and close relatives	Problematic to analyze except by cosegregation in families. As the evidence for linkage (by definition) will be scarce, the finding will be limited to most recognizable variants

Such MFDs as cardiovascular disorders, diabetes, cancer, and other late-onset diseases are generally more common at the post-reproductive age. Consequently, they are probably less subjected to selection. Traits with complex inheritance and late-onset diseases are more sensitive to changes in environmental conditions as a result of the accumulation of defects in homeostatic systems, but they exhibit higher variability than the traits directly related to reproductive success [9]. Although precise mechanisms maintaining variation of these traits are poorly studied, three main groups of them are known [10, 11].

First, there are variants deleterious both in the early life of the organism and later, which consequently are efficiently eliminated by natural selection, which explains their low frequencies in populations. The second category include variants selectively neutral in the early life, but with a negative effect at older age, which means that they are under weak selection and can have a high population frequency. The third group consists of variants that are advantageous in the early life but deleterious later and maintained by selection at an intermediate frequency. These theoretical considerations on the mechanisms of maintaining the genetic variation architecture were reflected in hypotheses (evolutionary models) of biological basis of the diseases with the onset in postreproductive period of the human development. These include the following hypotheses: CD/CV (common disease/common gene variant), MA (mutation accumulation) and AP (antagonistic pleiotropy), CD/RA (common disease/rare allele) and CD/FV (common disease/fixed variant). These hypotheses are described in detail in literature [9, 10, 12].

However, these models of genetic variability should be verified and alternative hypotheses should be tested. It is clear with respect to genetic variation of Mendelian forms of pathology that each mutation is rare, mutation diversity is high at each locus, and each mutation is necessary and sufficient for the formation of the disease phenotype. For MFDs, the genetic model postulates that affected individuals are those trespassing a biological risk threshold, carrying many "pathological" genes intricately imteracting with one another and with envrionmental factors. This model does not explain the existing phenotypic variability of multifactorial diseases. Moreover, new data (nonadditivity of gene interaction, "missing heritability", involvement of epigenetic mechanisms) indicating the importance of evolutionary context in considering genetic variability in human populations have been rapidly accumulating.

EVOLUTIONARY MEDICINE: NATURAL SELECTION AND GENETIC DRIFT

In evolutionary medicine, disease is considered as the result of a conflict between gene selection that have occurred in the past and the present-day requirements, the inability of our genes to match the modern lifestyle. Numerous human diseases were analyzed in this context (see Table 2). One of the earliest hypotheses in this line is related to the explanation of the increase in the proportion of individuals with excessive weight and diabetes in modern populations ("thrifty genes" hypothesis) [14]. Its author hypothesis suggested that the abundance of these diseases in contemporary human populations results from positive selection of alleles responsible for storing carbohydrates as fat, which is advantageous for survival during starvation. Such "thriftiness" is supposedly deleterious today, as deposition of fat upon endless consumption of high-calory food leads to obesity and insulin resistance. This article by James Neel [14] was republished after nearly 40 years in *Bulletin WHO* (1999, vol. 77, pp. 694–703), which testifies to the official acceptance and continuing popularity of the "thrifty genes" concept in the development of obesity and diabetes. In section "Some Eugenic Considerations" of this article, the author argues that, because of fluctuations in food supplies existing in many world regions, "the diabetic genotype should be preserved in the interests of the humankind." This means that Neel prefers the eugenic approach in controlling the genotype to rec-

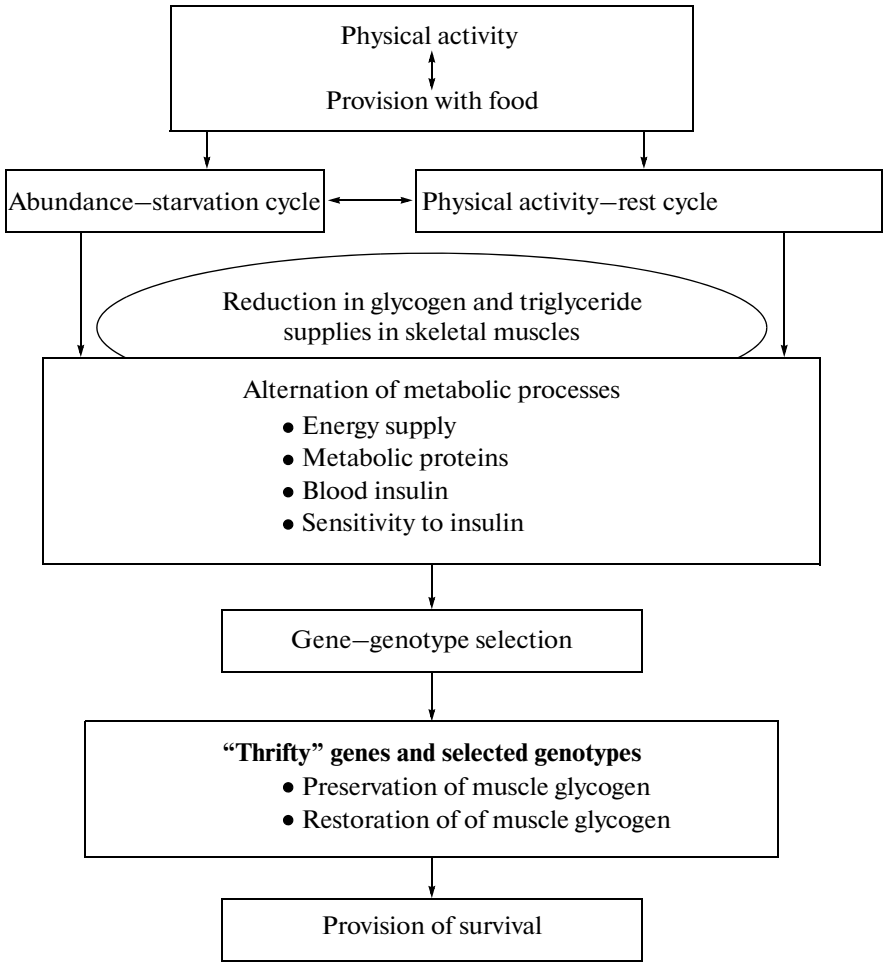
Table 2. Human diseases in the Evolutionary Perspective [13]

Disease/condition	Hypotheses
Sickle-cell anemia (SCA)	Individuals homozygous for the SCA mutation develop the disease but heterozygous carriers are more resistant to malaria, which maintains high frequencies of the mutant allele in populations
Diabetes mellitus	Mechanisms of food storage and use that were optimal and favored by selection in primitive populations of hunters–gatherers proved to be disadvantageous in the modern era of food excess. The "thrifty genotypes" concept
Food deficit	In the course of evolution, humans lost their ability to synthesize some vitally important substances (e.g., vitamina) and currently depend on obtaibnign them from environmental sources
Hypertension	"All-or-nothing" response to potentially dangerous stimuli maintained by selection in primitive human populations triggered physiological mechanisms that are potentially adverse in the contemporary society
	Evolutionarily optimal coordination or synchronization of growth and blood supply of target organs in postnatal development is disturbed due to modern tendency to overeating
AIDS	Human immunodeficiency virus causing AIDS stemmed from viruses that infected other species but later mechanisms of fast mutation were developed, particularly for drug resistance or avoiding immune system control
Cancer	Mechanisms of DNA repair and immune control normally controlling cell consition in the modern urbanized society become hostile on stimulation by toxins, diets, etc.
Myopia	Nearsigntedness leads to considerable selective disadaptation, but modern corrective lenses resulted in preservation of genetic variants predisposing to myopia, which led to general eyesight impairment in humans
Adverse response to drugs	Genes of drug metabolism play an important role in utilization of all xenobiotics in humas body; they were formed by coevolution of plants and animals, which reulted in synthesis of a wide variety of xenobiotics
Resistance to antiphthisic drugs	Antibiotics used against the agent of tuberculosis in fact promoted the evolution of the pathogen, which generated "escape mutants" that are resistant to modern antibiotics
Asthma	Primitive agrarian liestyle promoted the development of the functional lung properties that could react to and withstand mould and yeasts; in the modern urbanized environments with the air containing multiple substances and respiratory irritants (pollutants), these features are hyperreactive

ommendations on lifestyle, upbringing, development of dietary habits, etc. Incidentally, in the late 1980s, based on the results of his work with South African tribes, Neel admitted that this hypothesis was wrong and most data underlying it were currently refuted [15]. However, after a decade he stated that the "thrifty genotypes" hypothesis considered with regard to combination of type 2 diabetes mellitus (DMII), essential hypertension, and obesity would enter the 21st century as "syndromes of impaired genetic homeostasis" [16].

A new view on the Neel's opinion is presented in the hypothesis by M. Chakravarty and F. Booth [17]. This hypothesis states that in ancient hunters–gatherers periods of overeating and starvation alternated with obligatory periods of physical activity and rest. According to these authors, more than a thousand years ago *Homo sapiens* developed metabolic pathways that were adjusted to maintain and coordinate the overeating–starvation and physical activity–rest cycles. The modern humans have a different lifestyle, characterized by a constant excess of food and the absence of physical exercise, which contradicts the biochemical cycles programmed by evolution. The elimination of cyclicity of the corresponding metabolic processes ultimately results in metabolic disorders, namely obesity and DMII (figure).

Some authors critically comment the "thrifty genes" hypothesis. For instance, one of the criticisms is as follows [18]: periods of starvation sufficiently strong to increase mortality occur only once in several generations. Neolithic mortality was associated with infectious diseases rather than with body weight. Mortality was high mainly in young children and old individuals. Therefore, selection in favor of "thrifty" alle-



Scheme of the putative relationship between the physical activity cycles and metabolic processes that are controlled by "thrifty" genes and genotypes.

les would be very weak and would not play a significant role in spreading of these alleles. Incidentally, one of the highest selection coefficients in humans, giving the adaptive advantage of 5–10%, is known for the capability of assimilating milk products caused by persistent expression of genes *LCT* and *MCM6* [19, 20]. The "thrifty genes" hypothesis for DMII and obesity was not confirmed in relation to these diseases. Recently, Helgason et al. [21] showed that a haplotype of transcription factor gene *TCFL2* explains 20% of the DMII risk in Europeans and Africans, but the risk allele was associated with reduced rather than excessive weight in diabetes patients [21].

Speakman [18] calculated that if the "thrifty" genes underlying evolutionary beneficial obesity were under such long-term selection, we would have all inherited these genes and would have been all obese. This is obviously not true. Consequently, either obesity must be under balancing selection, or the alleles of corresponding genes are spread as a result of other processes. According to Speakman, this alternative process is gene drift: the mutant alleles in question are not selected for, being selectively neutral. This author termed these genes "drifty genes" and referred to the individuals who inherited predisposition to the disease as "losers in genetic lottery." "All evolutionists agree that gene drift cannot explain adaptive evolution. But evolution is not all adaptive" [22].

In any of the above hypotheses of chronic human disease development, a genetic component, differently included in the mechanism of the disease progress, can be traced. A special place among the "thrifty" hypotheses is occupied by the "thrifty phenotypes" hypothesis [23]. This hypothesis assumes the leading role of the environment in the embryonic development and growth, which ultimately determines the adult pathology. For DMII, the hypothesis suggests the following scenario. Low weight at birth can be retained during the first year of life, which is explained by various factors of the nutritional environment (mother's diet, her hormonal status during pregnancy, placental development). The embryo adapts to its environment, becomes "nutritionally thrifty," the weight of Langerhans islets in its pancreas reduces,

Table 3. SNPs in genes for MFS predisposition in the evolutionary context

Gene	SNP ID	Pathology/trait	P ($-\log_{10}$)
IL12RB2	rs3790564	Primary biliary liver cirrhosis	11.1
	rs3790567	Primary biliary liver cirrhosis	10.5–11.1
APOB	rs693	Chelesterol in low-density lipoproteids	10.5–16.4–21.0*
		Total cholesterol	22.0
LPP	rs1464510	Gluten enteropathy	8.3–39.5*
		Vitiligo	11.0
RNLS	rs10509540	Type 1 diabetes mellitus	28.0
FTO	rs6499640	Body weight index	12.4
		Weight	13.2
	rs1421085	Obesity	28.0
	rs1558902	Waist circumference	18.3
		Obesity (extreme)	12.2
	rs8050136	Type 2 diabetes mellitus	5.2–12.0–13.2–16.7*
		Body weight index	47
		Weight	35.3
	rs9939609	Body weight index	50.4
	rs9941349	Obesity (extreme)	11.2
NT5C2	rs11191548	Systolic blood pressure	23.2
TYR	rs1393350	Melanoma	13.7
PIGU	rs910873	Melanoma	15.0

* Estimates obtained in replicate studies.

and its growth decreases. The hormonal and metabolic status formed in this embryo is inadequate, and sometimes adverse, in the adult life, under conditions of normal and sufficient nutrition. The low insulin secretion and insulin resistance lead to DMII and obesity, sometimes mediating their combination with arterial hypertension and coronary disease.

Gibson [24] formulated the concept of decanalization as a specific class of genotype–environment interaction ("buffering") for explaining the origin of some MFDs (DMII, immune diseases, mental disorders) and their growth in the present-day populations. This concept includes three key notions: stabilizing selection, cryptic genetic variation, and dramatic cultural changes.

Millions of years of operation of stabilizing selection provided not only the optimal average level of important physiological functions (glucose metabolism, immunoreactivity, cognitive characteristics), but their strong stability (buffering, canalization). In this, the authors of the concept, according to their own statement, follow Schmalhausen and Waddington [25]: a population is canalized if most individuals of this population cluster around the optimal phenotype. Decanalization in the population is indicated by altering population risk of the disease examined: this risk should increase from a frequency of less than 1% in the ancestral populations to 10% in the modern community. The primary source of susceptibility to a chronic disease is cryptic genetic variation. This variation is revealed, released and nonlinearly interacts with the environmental factors, which have drastically changed in the last 100 years (diet, tobacco smoking, pollution, stress).

GENOMIC TARGETS FOR NATURAL SELECTION

The known number of genes subject to selection in human populations is over 2?000. However, both their number and spectrum are different as reported by different authors, while the degree of overlapping is low (less than 1/3) [26]. This situation may be explained by using different statistical methods (F_{ST} statistics, LD test, Tajima's D test, comparison of the synonymous/nonsynonymous substitution ratio (d_N/d_S), etc.) and methodological approaches for estimating selection effect (e.g., comparing geneic variation between human populations or between human and

Table 4. Number of genes with high F_{st} involved in various biological processes (from [34])

Biological processes	Races compared			Total
	Negroids–Caucasoids	Negroids–Mongoloids	Caucasoids–Mongoloids	
Apoptosis (147)	20 (13.6%)	32 (21.8%)	17 (11.6%)	46 (31.3%)
Blood circulation and gas exchange (13)	1 (7.7%)	2 (15.4%)	0 (0%)	2 (15.4%)
Blood coagulability (40)	7 (17.5%)	6 (15.0%)	5 (12.5%)	9 (22.5%)
Immune response (151)	20 (13.2%)	19 (12.6%)	10 (6.6%)	35 (23.2%)
Lipid metabolism (53)	11 (20.8%)	9 (17.0%)	5 (9.4%)	18 (34.0%)
Total (364)	51 (14.0%)	63 (17.3%)	32 (8.8%)	110 (30.2%)

chimpanzee populations) [26]. Depending on the analytical approaches chosen, the spectra of selected genes detected by different authors may mark various stages of human evolution, from the period of species divergenence to the present.

Among 50 genes for which strict selection effects were shown both at the early stages of human evolution (demonstrated by comparing SNPs between human and chimpanzee), and for modern populations, genes were recorded that are involved in the regulation of the following biological processes: olfaction (*OR2W1*, *OR511*, *OR2B2*, *C20orf185*), host–pathogen interaction (*CMRF35H*, *CD72*, *PTCRA*, *APOBEC3F*, *GZMH*), spematogenesis (*USP26*, *C15orf2*, *PEPP-2*, *TCP11*, *HYAL3*, *TSARG1*), oncogenesis (*HUAL3*, *DFFA*, *PEPP-2*, *C16orf3*, *MMP26*, *FLJ32965*), apoptosis (*PPP1R15A*, *HSJ001348*, *TSARG1*, *GZMH*) [27]. Akey et al. [28] assigned to candidates for selection 174 genes, of which 156 exhibited very high F_{st} values (inclusion criterion: for autosdomal genes minimum for one SNP $F_{st} \geq 0.45$; for X-chromosomal genes, $F_{st} \geq 0.45$), while for 18 genes, the F_{st} estimates were very low (inclusion criterion: two SNPs in gene with $F_{st} = 0$ and one SNP with $F_{st} \leq 0.005$). The products of these genes perform different functions (enzymes, transduction signals, transcription regulators, etc.) and are involved in various biological processes (intercellular interactions, cell growth, transcription, transport, development, etc.). Wang et al. [29] listed 112 genes with evident selection signatures, which were assigned to the following functional categories: reproduction (7%), host–pathogen interaction (10%), cell cycle (13%), protein metabolism (15%), neuron functioning (17%), and DNA metabolism, including putative transcriptional factors (21%). These results, as well as the evidence reported in other studies [30, 31], indicate that many genomic regions are subject to natural selection.

In spite of certain controversies in the information on selected genes reported by different authors, it seems important that for some of these genes (8.0–19.7% [28, 29]) significant associations with pathological states and traits were established in GWAS (Table 3). Selection signatures have been documented for a number of other candidate genes for multifactorial diseases: *ACE*, *AGT*, *CYP3A5*, *GNB3*, *GRK4*, *SCG2* (involved in regulation of arterial tension); *APOE*, *LDLR*, *PCSK9*, *SCARB1*, *ANGPTL4* (lipid metabolism); *CAPN10*, *TCF7L2* (energy metabolism); *F7*, *F9* (blood clotting); *IL4*, *IL4RA*, *IL13*, *IL1A*, *MMP3* (inflammation) [10].

The world population is characterized by high genetic heterogeneity in SNP frequencies of candidate genes for common diseases [10, 33]. In some cases, tenfold and greater interpopulation differences (e.g., for allele frequencies of genes controlling the lipid level) were reported for allele frequencies of 621 loci, whose genes showed association with complex human diseases and traits upon GWAS. The gene differentiation at these polymorphisms widely varies among disease groups (from $F_{st} = 0.067$ for attention deficit/hyperactivity syndrome to $F_{st} = 0.164$ for cardiovascular disease) and among SNPs within one pathology/trait (e.g., for growth-associated SNP $F_{st} = 0.008–0.504$; for those associated with the lipid level, $F_{st} = 0.006–0.520$) [33]. Similarity in allele frequencies is significantly low among populations from different continents (as shown for South Asia, Europe, Africa) than within these geographical regions. This concerted variation of allele frequencies in geographically close population groups may reflect effects of stabilizing selection.

Evidence for the effect of selection on the formation of genetic diversity was reported for candidate genes for cardiovascular diseases, involved in six biological processes: apoptosis, blood circulation and gas exchange, homeostasis, immune response, and lipid metabolism [34]. In spite of the fact that statitically significantly higher values of F_{st} than those of the "neutral" distribution were shown for a relatively low portion of SNPs (1.05–2.09%), the number of genes involved in different biological processes (excluding

Table 5. Genes under selection in human populations and putative selective factors

Gene (SNP)	Putative selection factor	Phenotype (disease, trait)	Source
CYP3A5 (*1/*3)	Climate (salt metabolism)	Hypertension, preeclampsia	[36]
AGT (M235T)	The same	Hypertension, preeclampsia	[36]
LEPR (R109K)	Climate (cold resistance)	Impaired metabolism	[37]
FABP2 (A54T)	The same	Impaired metabolism	[37]
GNB3 (C825T)	"	Arterial tension	[38]
FADS2 (rs174570)	Ecology	Lipid metabolism	[39]
TNXB (rs2269426)	Diet (meat and dairy products)	Number of eosinophils	[39]
ZCNQ1 (rs2722425)	Diet (tubers and root crops)	Fasting glucose	[39]
KCNQ1 (rs2237892)	The same	Type 2 diabetes mellitus	[39]
MTRR (rs162036)	"	Pholate metabolism	[39]
IL 4 (haplotypes)	Unknown pathogen	Asthma	[40]

homeostasis), for which F_{st} estimates were higher than "neutral" at least for one SNP, was significant (Table 4). The proportions of genes with high coefficient of gene differentiation were different among pairs of races compared. For instance, F_{st} values for a high fraction of genes regulating apoptosis, blood circulation, and gas exchange, was higher in the Negroid–Mongoloid pair than in comparisons of Caucasoids with Negroids and Mongoloids. According to the authors, one of the explanations of this fact may be local adaptation. Genes with local adaptation are thought to be as follows: in comparison of Negroids and Caucasoids, genes of immune response (*GRB2*, *IKBKB*), apoptosis (*ARHGEFI*, *RIPKI*), and lipid metabolism (*PMVK*); in comparison of Negroids and Mongoloids, genes of immune response (*GRB2*), apoptosis (*BCL2LI*), and lipid metabolism (*PMVK*); in comparison of Caucasoids and Mongoloids, genes of immune response and apoptosis (*IL4*, *IL6*) and genes of blood coagulation (*F2*).

The above studies were mainly focused on SNPs localized in coding gene regions. However, selection can also act on regulatory regions of the genome. High interethnic differences in allele frequencies in regulatory regions may determine specific profiles of gene expression, which was conclusively shown by Spielman et al. [35]. For instance, statistically significant differences in the gene expression between Caucasoid and Mongoloid populations were recorded for 26.1% of genes examined (including 35 genes showing greatest differences, $p = 6.32 \times 10^{-6}$–9.78×10^{-31}). Genes with maximum expression differences included *DPYSL2* ($p = 9.78 \times 10^{-31}$), *UGT2B17* ($p = 1.03 \times 10^{-18}$), *FCER2* ($p = 8.53 \times 10^{-15}$) (for these genes, the expression level was higher in Caucasoids); *CTSS* ($p = 7.52 \times 10^{-18}$), *SLC2A5* ($p = 1.24 \times 10^{-18}$), *NFIL3* ($p = 5.85 \times 10^{-17}$), *DNAJB9* ($p = 9.11 \times 10^{-15}$) (higher in Mongoloids). Based on the data on the expression of 1097 genes, all individuals examined with practically 100% probability clustered in two groups: "Mongoloids" and "Caucasoids." The authors of the study demonstrated that 18–81% of the total variation in the gene expression are explained by SNP frequencies in the populations studied. For example, differences in the gene expression between the two groups compared for gene *UGT2B17* were recorded at $p < 10.3 \times 10^{-18}$; the frequency of allele *A* at rs3100645, associated with higher expression level, was 0.68 in Caucasoids and 0.15 in Mongoloids [35].

Since the expression level is largely determined by genetic characteristics of regulatory factors, investigation of SNPs in regulatory genome regions can also shed light on understanding differences in disease susceptibility among populations. In this connection, it seems interesting that population subdivision at SNPs localized in coding regions is lower ($F_{st} = 0.107$) than that at SNPs localized in introns ($F_{st} = 0.118$) and other noncoding genome regions ($F_{st} = 0.123$), which may reflect stronger selection on regulatory genome parts. Wang et al. [29] showed that 35% of SNPs with documented dselection signatures were localized at a distance of at least 100 kb from known genes. At that, the gene differentiation at individual SNPs can vary in a wide range, as, for instance, was shown for polymorphic variants associated with common diseases according to GWAS [33].

Identifying specific environmental factors that determine genetic features of the gene pool structure in particular populations is a more complicated task, but such evidence has been accumulating (Table 5). In their recent study, Hancock et al. [39] detected a number of SNPs (predominantly those localized in genes and particularly nonsynonymous) that showed correlation with environmental parameters, life style (subsistance), and diet. The authors also presented explanations of the associations of genetic markers with the environmental parameters. For instance, for diet they found specific SNPs in the genes involved in metabolism of starch and sucrose (*GAA*, *GBE1*, *GBA3*), folate biosynthesis (*MTRR*) for populations in which the diet was based on

tubers and root crops; *PLRP2* (hydrolyzes galactolipids, the component of triglycerides in plants), if the diet was based on cereals. A strict association of such ecological parameter as "polar region" was established for SNPs localized in genes whose products are involved in energy metabolism (*ME2*, *ME3*). For a number of SNPs associated with the ecological parameters and diet characteristics examined, genome-wide association studies showed highly significant association with diseases or pathogenetic traits. For instance, rs17450 localized in gene *FADS2*, showed association with the ecological region (humid tropics, $p = 2.00 \times 10^{-5}$) and lipid metabolism parameters ($p = 2.00 \times 10^{-10}$–4.00×10^{-13}); rs2237892 in gene *KCNQ1* showed associaiton with diet features (cereals as the main diet component, $p = 1.49 \times 10^{-4}$) and type 1 diabetes mellitus ($p = 1.70 \times 10^{-42}$), etc. [39].

Fitness of individuals or populations is determined by combination of genotypes at nonlinked loci rather than by distinct genes (alleles) or blocks of linked genes. Selection can act on pairs of nonlinked, but coevolving in concert genes [41] or on combinations of genotypes at several loci that are unfavorable for the populations. Ultimately, selection is targeted at maintaining the optimal genetic diversity of the population [5]. Irrespective of the marker systems used, estimates of genetic differentiation in human populations at a set of loci show surprising stability and F_{st} values in them rarely exceed 0.15 [42, 43]. It is important that the effects of selection may have different direction at different ontogenetic stages, as was demonstrated for such integral parameter as individual heterozygosity [44].

ONTOGENY OF CHRONIC HUMAN DISEASES

Although the role of natural selection in the evolution of organisms, including the formation of hereditary predisposition to human chronic multifactorial diseases, is generally accepted, logical extrapolation of the evolutionary idea on possible significance of selective processes for tissues or cells in the pathogenesis of these diseases is still not evident and remains under dispute. Since this involves possibility of selection of somatic cells carrying mutations during an individual lifetime, the participation of this phenomenon in disease pathogenesis was termed disease ontogeny [45]. In any case, somatic mutations are random events, but, as justly noted by the evolutionary biologist Ayala in his comments on evolution and natural selection, randomness is an essential part of the evolutionary process [46].

In oncology, the mutation hypothesis of carcinogenesis prevails, which states that all tumors are mutant phenotypes. For instance, according to the "two-hit" hypothesis of carcinogenesis by Knudson (1971), the individual that inherited from one of his parents a mutation (point mutation or microdeletion) in a tumor-suppressor gene, during his lifetime acquires an additional somatic mutation in the normal allele of the same gene, which results in loss of heterozygosity (LOH) causing the complete absence of the tumor-suppressor gene in the cell [47]. Two such mutations may also arise in one locus of the cell genome in individuals lacking germline mutations. This mechanism underlies the formation of many human tumors (retinoblastoma; lung, breast, bladder, liver cancer; osteosarcoma; Wilms tumor; etc.).

However, the mutation component of carcinogeneisis is not the sole mechanism of cancer development. It was postulated that along with mutation, cancerogenesis additionally involves selection of mutant cells in the corresponding tissues, thus presenting an example of "Darwinian medicine" [45, 48]. These theoretical views have been currently supported by experimental evidence. Vineis and Berwick (2006) presented data on selection of cells carrying premutations, while specific intratissue environmental factors of selection were termed *selectogenes* [48].

Developing this view on carcinogenesis, Gottlieb et al. [45] extended this hypothesis, applying their original data on prostate cancer genetics to ontogeny of other MFDs. These authors showed that prostate tissue cells of cancer patients carry multiple variants of the androgen receptor gene *AR* containing different numbers of CAG repeats. However, a similar picture was recorded in the prostate tissues of healthy individuals, including those of a one-year-old child. Cells of the healthy tissue exhibit intra-tissue genetic heterogeneity (ITGH) with minor and major variants. Since the length of the CAG repeats is directly related to the AR protein activity, the androgen levle in the gland tissue (microenvironment) acts as a factor of selection of cells carrying either minor, or major genetic variants. The oncological phenotype is associated with the major CAG repeat variant of the *AR* receptor gene. The authors of this hypothesis suggest that minor variants in healthy tissues are subjected to selection when intra-tissue and inter-tissue condictions change. Such evolutionary events can be present in the ontogeny of not only cancer but also other diseases (cardiovascular diseases and diabetes).

Happle (1986) described the phenomenon of paradominant inheritance [49], which McKusick called "an intriguing hypothesis" [50]. In essence, paradominant inheritance is a combination of inherited genetic alterations with somatic disturbance of the same genes: mutations are transmitted over generations in heterozygous state because their carriers lack phenotypic manifestations and the homozygous state is lethal. The character is expressed only in case of LOH in some somatic cells. The main distinction between this LOH form and other forms, associated with diseases, in particular, with malignant tumors, is that in the former case, LOH occurs in early ontogeny (variation of postzygotic LOH), resulting in a new cell clone, which, in its turn, produces a mosaic tissue region. Recently, Limaye et al. [51] reported evidence showing the inportance of local structural distur-

bances (somatic mosaicism) that, along with combination of inherites same gene variants, cause many vascular anomalies, appearing during vasculogenesis, angiogenesis, and lymphoangiogenesis. The author of the term "paradominant inheritance", commenting on this study, emphasized that this phenomenon belongs to the category of non-Mendelian inheritance and differs from other LOH cases in ontogeny of diseases, for instance, retinoblastoma representing a classical variant of autosomal dominant inheritance [52]. Apparently, such LOH-associated diseases do not need evolutionary hypotyheses to explain the mechanisms of formation of the pathological phenotype.

Mechanisms of ontogenetic development and disease ontogeny are determined not only by the genome–phenome interaction, but also by the effects of epigenetic inheritance. Epigenetic variation (CpG dinucleotide methylation in DNA, histone covalent modifications, activity of small regulatory and interfering RNAs) in MFD pathogenesis has not been sufficiently quantified, but some landmarks were determined. For instance, during ontogenesis, the loss of epigenetic control of gene suppression in normal tissues occur at a rate that is one or two orders of magnitude higher than the DNA somatic mutation rate [53]. It is thought that the rate of epimutaion exceed the gene mutation rate by the same value (one to two orders of magnitude) [54]. The contribution of methylated genes controlling development and morphogenesis (homeobox cluster genes) measured by the odds ratio, varied from 3.6 to 43.3, which is substantially higher than that observed for structural gene variants [55].

Novel technological tools for analysis of genetic and epigenetic variability open perspectives for obtaining new information related to ontogeny of chronic human diseases. These tools include massive parallel sequencing [56] and massive evaluation of epigenetic information using microarray techniques, which simulataneously analyze methylation state of individual CpG sites of many genes [57].

CONCLUSIONS

Evolutionary knowledge in understanding the human nature, regularities of health preservation and pathogenesis build the foundation and form discussion areas that must bring medical science and practice closer to the impatiently awaited personalized medicine and testing hereditary predisposition to diseases, the unified theory of human diseases and their new classification that does not coincide with the traditional nosological principle. Recall Huxley's warning that without the evolutionary approach and understanding human evolution, our picture of the world would be false and our attemts to transform civilization may fail [58].

Let us summarize our discussion of evolutionary ontogenetic issues of human pathology in several points.

1. Modern genomic studies provided evolutionary medicine with extensive information on the structure of the human genome and the genomes of pathogens causing infectiuos parasitory diseases. For MFD genetics, the research results are still modest: according to GWAS, about 800 SNPs assigned to susceptibility gene alleles were identified for 150 chronic diseases and traits [59]. The total SNP proportion in the structure of hereditary component of MFD susceptibility does not exceed 10–12%, i.e., the problem of "missing heritability" for this category of diseases is to be solved. The advances in the development of the "-omic" technologies and the evolutionary context of considering their results give hope to resolve these issues.

2. The widely clinically known enigmatic phenomenon of comorbidity (combination of several diseases in one individual) is a consequence of genetic adaptation of individuals and populations to changing environments. However, not only combinations of different diseases in individuals and their closest relatives (syntropies), but exclusions of such "encounters" (dystropies) are a natural phenomenon that has an evolutionary basis. Our explanation of this phenomenon from the viewpoint of genetics stated in the 1980s [60] today corresponds to the modern concepts of diseasom [61] and network medicine, implying that MFD pathogenetics involves replacement of hierarchic networks by self-organizing new networks [4, 62].

3. It was noted that the 1980s were characterized by complete neglect of ontogenetic issues by the Modern Synthesis [63]. However, exactly at that time, in 1986, the "two-hit" model of carcinogenesis was confirmed [47], according to which the disease develops in individuals who inherited the corresponding mutation from one of the parents and acquired another one in the same locus during the ontogeny. The mechanism of disease development that was proven for cancer is seen in other pathologies. Such phenomena as somatic mosaicism, paradominant inheritance, loss of heterozygosity, and epigenetic variability have been increasingly actively studied in pathogenetics of chronic human diseases. Concurrent investigation of inherited genetic events interacting with those acquired in individual development seems particularly important and promising.

4. Evolutionary studies of chronic human diseases shed light on general regularities of their appearance and development. However, the uniqueness of each individual imparts significant modifications to the clinical polymorphism, on which individual therapy, prevention, and prognostics are based. In this sense, the statement that "each disease has its own genetic architecture that depends on the human evolutionary history" is true. Medical science has only just began to accumulate information on individual human

genomes in the clinical context, which manifests the beginning of new perspectives and exciting research.

ACKNOWELEDGMENTS

This study was supported by the Federal Targeted Program "Scientific and Pedagogical Personnel of Innovative Russia", state contract no. 02.740.11.0284, and Federal Targeted Program "Personnel", theme "Investigation of Pleiotropic Gene Effects at Different Functional States of Human Organism," state contract no. P-1288, Aug. 31, 2009.

REFERENCES

1. Green, E.D. and Guyer, M.S., National Human Genome Research Institute: Charting a Course for Genomic Medicine from Base Pairs to Bedside, *Nature*, 2011, vol. 470, pp. 204–212.
2. Di Rinzo, A. and Hundson, R.R., An Evolutionary Framework for Common Diseases: The Ancestral-Susceptibility Model, *Trends Genet.*, 2005, vol. 21, no. 11, pp. 596–601.
3. Nesse, R.M., Bergstrom, C.T., Ellison, P.T., et al., Making Evolutionary Biology a Basic Science for Medicine, *Proc. Natl. Acad. Sci. USA*, 2010, vol. 107, suppl. 1, pp. 1800–1807.
4. Olsson, L., Levit, G.S., and Hossfeld, U., Evolutionary Developmental Biology: Its Conceptions and History with a Focus on Russian and German Contribution, *Naturwissenschaften*, 2010, vol. 97, pp. 951–969.
5. Altukhov, Yu.P., *Geneticheskie protsessy v populyatsiyakh* (Genetic Processes in Populations), Moscow: Akademkniga, 2003.
6. Altukhov, Yu.P., Korochkin, L.I., and Rychkov, Yu.G., Hereditary Biochemical Diversity in the Processes of Evolution and Ontogenesis, *Russ. J. Genet.*, 1996, vol. 32, no. 11, pp. 1256–1275.
7. Wray, N.R., Goddard, M.E., and Visscher, P.M., Prediction of Individual Genetic Risk of Complex Disease, *Curr. Opin. Genet. Dev.*, 2008, vol. 18, pp. 257–263.
8. Cirulli, T.T. and Goldstein, D.B., Uncovering the Roles of Rare Variants in Common Diseases through Whole-Genome Sequencing, *Nat. Rev. Genet.*, 2010, no. 11, pp. 415–426.
9. Wright, A., Charlesworth, B., Rudan, J., et al., A Polygenic Basis for Late-Onset Disease, *Trends Genet.*, 2003, vol. 19, no. 2, pp. 97–106.
10. Ding, K. and Kullo, I.J., Evolutionary Genetics of Coronary Heart Disease, *Circulation*, 2009, vol. 119, pp. 459–467.
11. Barton, N.H. and Keightley, P.D., Understanding Quantitative Genetic Variation, *Nat. Rev. Genet.*, 2002, no. 3, pp. 11–21.
12. Puzyrev, V.P., Genetics of Multifactorial Diseases: Between Past and and Future, *Med. Genet.*, 2003, vol. 2, no. 12, pp. 498–508.
13. Schork, N.J., Cardon, L.R., and Xu, X., The Future of Genetic Epidemiology, *Trends Genet.*, 1998, vol. 14, pp. 266–271.
14. Neel, J.V., Diabetes Mellitus a "Thrifty" Genotype Rendered Detrimental by "Progress"?, *Am. J. Hum. Genet.*, 1962, vol. 14, pp. 352–353.
15. Neel, J.V., Update to Study of Natural Selection in Primitive and Civilized Human Populations, *Hum. Biol.*, 1989, vol. 61, pp. 811–823.
16. Neel, J.V., Weder, A.B., and Julius, S., Type II Diabetes, Essential Hypertension and Obesity as "Syndromes of Impaired Genetic Homeostasis": "Thrifty" Genotype Hypothesis Enters the 21st Century, *Perspect. Biol. Med.,* 1998, vol. 42, no. 1, pp. 44–74.
17. Chakravarty, M.V. and Booth, F.W., Eating, Exercise and "Thrifty" Genotypes: Connecting the Dots toward an Evolutionary Understanding of Modern Chronic Diseases, *J. Appl. Physiol.*, 2004, vol. 96, pp. 3–10.
18. Speakman, J.R., Thrifty Genes for Obesity and Metabolic Syndrome—Time to Call off the Search?, *Diab. Vasc. Dis. Res.*, 2006, no. 3, pp. 7–11.
19. Swallow, D.M., Genetic of Lactase Persistence and Lactose Intolerance, *Annu. Rev. Genet.*, 2003, vol. 37, pp. 197–219.
20. Tishkoff, S.A., Reed, F.A., Ranciaro, A., et al., Convergent Adaptation of Human Lactase Persistence in Africa and Europe, *Nat. Genet.*, 2007, vol. 39, pp. 31–40.
21. Helgason, A., Palsson, S., Thorleifsson, G., et al., Refining the Impact of *TCF7L2* Gene Variants on Type 2 Diabetes and Adaptive Evolution, *Nat. Genet.*, 2007, vol. 39, pp. 218–225.
22. Dawkins, R. and Coyne, J., *On Side Can be Wrong*, www.guardian.co.uk/science/2005/sep/01/schools.research/
23. Hales, C.N., Barker, D.J., Clark, P.M., et al., Fetal and Infant Growth and Impaired Glucose Tolerance at Age 64 Years, *BMJ*, 1991, vol. 303, pp. 1019–1022.
24. Gibson, G., Decanalisation and the Origin of Complex Disease, *Nat. Rev. Genet.*, 2009, vol. 10, pp. 134–140.
25. Gibson, G. and Wagner, G., Canalization in Evolutionary Genetics: A Stabilizing Theory?, *BioEssays*, 2000, vol. 22, pp. 372–380.
26. Biswas, S. and Akey, J.M., Genomic Insights into Positive Selection, *Trends Genet.*, 2006, vol. 22, no. 8, pp. 437–446.
27. Nielsen, R., Bustamante, C., Clark, A.G., et al., A Scan for Positively Selected Genes in the Genomes of Human and Chimpanzees, *PLoS Biol.,* 2005, vol. 3, no. 6, pp. 976–985.
28. Akey, J.M., Zhang, G., Zhang, K., et al., Interrogation a High-Density SNP Map for Signature Natural Selection, *Genome Res.*, 2002, vol. 12, pp. 1805–1814.
29. Wang, E.T., Kodama, G., Baldy, P., and Moyzis, R.K., Global Landscape of Resent Inferred Darwinian Selection for Homo Sapiens, *Proc. Natl. Acad. Sci. USA,* 2006, vol. 103, no. 1, pp. 135–140.
30. Weir, B.S., Cardon, L.R., Anderson, A.D., et al., Measures of Human Population Structure Show Heterogeneity among Genomic Regions, *Genome Res.,* 2005, vol. 15, pp. 1468–1476.
31. Carlson, C.S., Thomas, D.J., Eberle, M.A., et al., Genomic Regions Exhibiting Positive Selection Identi-

fied from Dense Genotype Data, *Genome Res.*, 2005, vol. 15, pp. 1553–1565.

32. http://www.ncbi.nlm.nih.gov/gene/

33. Adeyemo, A. and Rotimi, C., Genetic Associated with Complex Human Diseases Show Wide Variation across Multiple Population, *Public Health Genomics*, 2010, vol. 13, pp. 72–79.

34. Kullo, I.J. and Ding, K., Patterns of Differentiation of Candidate Genes for Cardiovascular Disease, *BMC Genet.*, 2007, vol. 8, p. 48, http://www.biomedcentral.com.1471-2156/8/48.

35. Spielman, R.S., Bastone, L.A., Burdick, J.T., et al., Common Genetic Variants Account for Differences in Gene Expression among Ethnic Groups, *Nat. Genet.*, 2007, vol. 39, no. 2, pp. 226–231.

36. Thompson, E.E., Kuttab-Boulos, H., Witonsky, D., et al., *CYP3A* Variation and the Evolution of Salt-Sensitivity Variants, *Am. J. Hum. Genet.*, 2004, vol. 75, no. 6, pp. 1059–1069.

37. Hancock, A.M., Witonsky, D.B., Gordon, A.S., et al., Adaptations to Climate in Candidate Genes for Common Metabolic Disorders, *PLoS Genet.*, 2008, vol. 4, p. e32.

38. Young, J.H., Chang, Y.P., Kim, J.D., et al., Differential Susceptibility to Hypertension Is Due to Selection during the out-of-Africa Expansion, *PLoS Genet.*, 2005, vol. 1, no. 6, p. e82.

39. Hancock, A.M., Witonsky, D.B., Ehler, E., et al., Human Adaptation to Diet, Subsistence, and Ecoregion Are Due to Subtle Shifts in Allele Frequency, *Proc. Natl. Acad. Sci. USA*, 2010, vol. 107, suppl. 2, pp. 8924–8930.

40. Sakagami, T., Witherspoon, D.J., Nakajima, T., et al., Local Adaptation and Population Differentiation at the Interleukin 13 and Interleukin 4 Loci, *Genes Immunity,* 2004, vol. 5, pp. 389–397.

41. Rohlfs, R.V., Swanson, W.J., and Weir, B.S., Detecting Coevolution through Allelic Association between Physically Unlinked Loci, *Am. J. Hum. Genet.*, 2010, vol. 86, pp. 674–685.

42. Kucher, A.N., Babushkina, N.P., Markova, V.V., et al., Variation in the Polymorphic Candidate Genes for Cardiovascular Disease among Four Ethnic Groups in Siberian Region, *Med. Genet.,* 2010, no. 5, pp. 24–34.

43. Stepanov, V.A., Genomes, Populations, Diseases: Ethnic Genomics and Personified Medicine, *Asta Nat.*, vol. 2, no. 4(7), pp. 18–34.

44. Altukhov, Yu.P., Allozyme Heterozygosity, Sexual Maturation Rate, and Longevity, *Russ. J. Genet.*, 1998, vol. 34, no. 7, pp. 751–760.

45. Gottlieb, B., Beitel, L.K., Alvarado, C., and Trifiro, M.A., Selection and Mutation in the "New" Genetics: An Emerging Hypothesis, *Hum. Genet.*, 2010, vol. 127, pp. 491–501.

46. Ayala, F.J., On Hundred Years without Darwin Are Enough!, *Genome Res.*, 2009, vol. 19, pp. 693–699.

47. Knudson, E. and Knudson, K., Retinoblastoma Tumor Suppressor: Where Cancer Meet the Cell Cycle, *Exp. Biol. Med.,* 2006, vol. 231, pp. 1271–1281.

48. Vineis, P. and Berwick, M., The Population Dynamics of Cancer: A Darwinian Perspective, *Int. J. Epidemiol.,* 2006, vol. 35, pp. 1151–1159.

49. Happle, R., The McCune–Albright Syndrome: A Lethal Gene Serving by Mosaicism, *Clin. Genet.*, 1986, vol. 29, pp. 321–324.

50. McKusick, V.A., *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetics Disorders*, Baltimore: The John Hopkins University Press, 1994, 11th ed.

51. Limaye, N., Boon, L.M., and Vikkula, M., From Germline to Somatic Mutations in Pathophysiology of Vascular Anomalies, *Hum. Mol. Genet.,* 2009, vol. 18, no. R1, pp. R65–R74.

52. Happle, R., What Is Paradominant Inheritance?, *J. Med. Genet.*, 2009, vol. 46, p. 648.

53. Bennet-Baker, P.E., Wilkowski, J., and Burke, D.T., Age-Associated Activation of Epigenetically Repressed Genes in Mouse, *Genetics,* 2003, vol. 165, pp. 2055–2062.

54. Horsthemke, B., Epimutations in Human Disease, *Curr. Top. Microbiol. Immunol.*, 2006, vol. 310, pp. 45–59.

55. Feinberg, A.P., Irizarry, R.A., Fradin, D., et al., Personalized Epigenomic Signatures That Are Stable over Time and Covary with Body Mass Index, *Sci. Translat. Med.*, 2010, vol. 2, no. 49, pp. 49–67.

56. Tucker, T., Marra, M., and Friedman, J.M., Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine, *Am. J. Hum. Genet.*, 2009, vol. 85, no. 8, pp. 142–154.

57. Laird, P.W., Principles and Challenges of Genome-Wide DNA Methylation Analyses, *Nat. Rev. Genet.*, 2010, vol. 11, no. 3, pp. 191–203.

58. Huxley, J.S., *The Uniqueness of Man*, London: Chatto and Windus, 1941.

59. Manolio, T.A., Genomewide Association Studies and Assessment of the Risk of Disease, *New Eng. J. Med.*, 2010, vol. 363, pp. 166–176.

60. Puzyrev, V.P. and Freidin, M.B., Genetic Sight on the Phenomenon of Combined Human Diseases, *Acta Nat.,* 2009, no. 3, pp. 57–63.

61. Goh, K.-J., Cusick, M.E., Valle, D., et al., The Human Disease Network, *Proc. Natl. Acad. Sci. USA*, 2007, vol. 104, pp. 8685–8690.

62. Barabasi, A., Gulbahce, N., and Localzo, J., Network Medicine: A Network-Based Approach to Human Disease, *Nat. Rev. Genet.*, 2011, vol. 12, no. 1, pp. 56–68.

63. Mueller, G.B., In Innovation Triad: An EvoDevo Agende, *J. Exp. Zool. (Mol. Dev. Evol.),* 2005, vol. 304B, pp. 487–503.

64. Weiss, K.M. and Terwilliger, J.D., How Many Diseases Does It Take to Map a Gene with SNPs?, *Nat. Genet.*, 2000, vol. 26, pp. 151–157.