REVIEW ARTICLES

Genetic background of phenotypic variation

Zsolt Boldogköi *

(Department of Biology, Faculty of Medicine, University of Szeged, Somogyi Bela st. 4, H-6720 Szeged, Hungary)

Accepted on May 25, 2007

Abstract A noteworthy feature of the living world is its bewildering variability. A key issue in several biological disciplines is the achievement of an understanding of the hereditary basis of this variability. Two opposing, but not necessarily irreconcilable conceptions attempt to explain the underlying mechanism. The gene function paradigm postulates that phenotypic variance is generated by the polymorphism in the coding sequences of genes. However, comparisons of a great number of homologous gene and protein sequences have revealed that they predominantly remained functionally conserved even across distantly related phylogenic taxa. Alternatively, the gene regulation paradigm assumes that differences in the *cis*-regulatory region of genes do account for phenotype variation within species. An extension of this latter concept is that phenotypic variability is generated by the polymorphism in the overall gene expression profiles of gene networks. In other words, the activity of a particular gene is a system property determined both by the *cis*-regulatory sequences of the given genes and by the other genes of a gene network, whose expressions vary among individuals, too. Novel proponents of gene function paradigm claim that functional genetic variance within the coding sequences of regulatory genes is critical for the generation of morphological polymorphism. Note, however, that these developmental genes play direct regulatory roles in the control of gene expression.

Keywords: evolution, gene network, gene regulation, genetic variance, phenotype variation.

Diversity, the fundamental hallmark of living systems, is manifested at all levels of biological hierarchy, including nucleic acids, proteins, biochemical and physiological processes, morphology and behavior. The animal body plan is specified by the genetic material in a very rigid manner that is, morphogenesis in most cases is highly resistant to environmental disturbance. Consequently, the polymorphism of morphological traits must be encoded by likewise diverse genetic material. The major question that arises is the origin of this genetic variation. The answer to this leads beyond our mere curiosity concerning an understanding of the principles of genetic organization. Deciphering of the algorithm underlying the operation of genetic material will be, among others, a central issue of future medicine. Insight into the genetic basis of phenotypic variation among individuals will provide us with a better understanding of complex diseases and a guiding principle for therapy design, which will be especially important in individual-centered healthcare in the not too far future. The theoretical clarification of the mechanism of evolution is strictly linked to this issue. Natural selection acts on traits and behaviors, while evolutionary changes proceed in DNA sequences. To resolve this apparent paradox, we have to decode the mechanism whereby the genetic blueprint controls the development and operation of phenotypic characters. It was earlier believed that phenotypic polymorphism was produced by the variance in the coding regions of genes, implying that variations exist in the effectiveness or functions of proteins. However, it was shown that protein function is highly conserved even across large evolutionary distances[1]. In the past few years a broad consensus has emerged that it is not the functions of proteins, but their expressions that vary among individuals in a population. This *cis*-regulatory polymorphism model holds that phenotypic variance is generated by the polymorphism in the controlling regions of genes. This hypothesis is based on the observation that genes exhibit extensive intraspecific allelic variations in their regulatory sequences[2,3]. Alternatively, the *trans*-regulatory polymorphism model hypothesizes that phenotypic variation is generated by the variance in the coding regions of DNA-binding protein genes (developmental transcription factor genes[4]) and protein-binding protein genes (components of cellular signaling pathways[5]). Further, the recently discovered regulatory RNAs are supposed to play a not yet ascertained role as *trans*-acting factors in gene

* E-mail: boldog@sb4.szote.u-szeged.hu

regulation. Thus, it appears that variance in the genetic regulatory machinery, rather than in the function of physiological genes, underlies phenotypic variation. The question is the relative extents to which these two regulatory mechanisms contribute to the anatomical and physiological polymorphism of organisms. A further subject of debate is whether intraspecific phenotypic polymorphism can be generated by single gene variants (or equivalently, by the variance in the simple sums of individual gene effects in the case of polygenic traits in population genetics), or whether alternatively, it is produced by the variation in gene networks (GNs). The classic Mendelian-Morganian view holds that polymorphism in a single gene alone can account for the variance of a certain phenotypic character. According to another view, phenotypic variance is generated by the polymorphism in GNs, which is mainly encoded in the *cis*-regulatory regions (CRSs) of GN components. According to the gene network-based concepts, the expression of a particular gene is a system property in that it is determined not merely by the regulatory architecture of this gene, but also by the expression profiles of other members of a GN whose regulatory sequences display variation, too. Coding variance in transcription factors further increases gene expression polymorphism.

1 The genetic architecture of phenotype

For an understanding of how genes produce phenotypic variance in a population, we first have to understand the machinery whereby the phenotype itself is determined.

1.1 One gene, one phenotype or one piece of phenotype—the view of the 20th century genetics

The brilliant experiments by Gregor Mendel shed light on some important aspects of the architecture and inheritance of genetic material. Mendel demonstrated that the hereditary material is comprised of discrete entities (later called genes); each trait in an individual is controlled by two hereditary factors, which can be either identical (homozygous) or different (heterozygous); depending on their relationship one variant of a gene (allele) can be dominant over another variant (s) or recessive. Mendel utilized a pea system, where each allelic variant could be unambiguously correlated to a particular phenotypic character. This type of inheritance was later conveyed as monogenic inheritance. The genetic determination of quantitative traits (such as height or weight) used the same explanatory scheme; these traits were regarded as the sums of individual gene effects (polygenic inheritance). Conversely, it was assumed that each gene made a specific, quantifiable contribution to a particular quantitative trait, and that genes exerted their effects independently of each other. In other words, in this Mendelian world it was considered that an allele was not only a necessary, but also a sufficient requirement for the determination of a phenotype or a piece of phenotype. The work on model organisms including fruit fly, viruses and bacteria that gave birth to molecular biology strengthened the earlier dogma of one gene, one phenotype of the early Mendelians. In recent decades, molecular geneticists have contributed to this misconception, e.g. through the ways they named their genes on the basis on the phenotypic effects they cause if mutated (especially in the case of *Drosophila*). This semantic imprecision has had an unfortunate effect on the perception of gene action: it has contributed to the belief that the invariance observed between the mutation of a gene and the resulting phenotype can be extrapolated as though the function of a gene is to ensure the normal function abolished by the mutation; and that phenotype is due to the exclusive function of a particular gene. Although, this over-simplistic view is apparently contradictory to common sense, it served for a long time not only as the research paradigm, but also as the conceptual basis of modern genetics and evolutionary biology. Even today, the aim in a considerable proportion of genetic investigations is to find a correspondence between a phenotype and a particular gene. Two basic approaches can be distinguished: In forward genetic research random mutations induced in the genome of a model organism are followed by attempts to identify the mutant gene accounting for the altered phenotype. In contrast, reverse genetics seeks to explore the function of a single gene by deleting it or by altering its expression, with a subsequent search for the phenotype obtained. In gain-of-function models of reverse genetics, a transgene is over-expressed by inserting it into an ectopic genomic location; in loss-of-function approaches, an endogenous gene is knocked out via gene targeting, or its expression is down-regulated by antisense blocking or is knocked down by RNA interference. The stereotypic conclusion of these studies is that the particular gene is responsible for the normal development of the phenotype altered by the genetic manipulation, which is the typical reductionist stance. Nevertheless, as yet no case has been demonstrated where the lines of causali-

ty could be mapped from a single gene to a phenotypic character. Molecular biological investigations, however, have produced a huge number of data on the pleiotropic effects exerted by the modified gene function on the operation of other genes and on epitasis, the multifactorial determination of a trait or behavior[6—8]. In the interpretation framework of reverse genetics, pleiotropy disturbs the pure effect of an examined gene on the phenotype presumed to be determined by this gene. This interpretation is in most cases apparently erroneous. The extent of error depends on how far the examined phenotype is from the gene product in the hierarchy of regulation. In the case of blood groups, the stringent correspondence between gene and the phenotype is acceptable since blood groups are distinguished on the basis of protein variants, which are the products of genes. Similarly, if the phenotype is a direct biochemical function of a protein, the direct relationship between genes and phenotype could be correct. However, if we make a great jump in the regulatory hierarchy and correlate a gene to, for example, a higher-order mental process such as depression[9], or speech[10], we fall into the trap of oversimplification in the interpretation of the results obtained. Ablation of a gene function can cause rearrangement of the expression profiles of other genes to such an extent that it is impossible to unravel the contribution of the given gene to a particular phenotype from the muddled web of causes and effects; which is due not only to the complexity, but also to the epistatic and pleiotropic effects of interacting genes. Alternatively, the phenotypic effect of a malfunctioned gene can be masked by compensatory mechanisms exerted by other genes (robustness). The major problem with the investigation attitude of reductionist genetics is that it restricts its focus to the effects of gene deletion on the structure and function presupposed to be altered, and neglects the interrelatedness of this gene with other systems. This view has led to adverse long-term consequences in the research paradigm of modern genetics by generating a false focus of investigation and improper interpretations of obtained results. Of course, most geneticists were aware of the importance of the interactive nature of genes; however, they neglected this situation for several reasons, including the need for a simplistic explanatory framework and the unavailability of suitable techniques for the investigation of intricate genetic interactions.

1.2 Interacting genes specify the phenotypes—the view of the post-genome era

It was developmental biologists who first explicitly declared that life is more complicated than a 1∶1 relationship between genotype and phenotype. Instead, genes create products that interact with one another and the environment in complex, hierarchical ways, to establish phenotypes. The result of genetic interplay is the continuous remodeling of gene expression profiles during development (e.g., in Ref. [11]) and in the adult body. One gene affects several phenotypic characters (pleiotropy), and the reverse is also true: a phenotype is determined by several genes (epistasis). From another aspect, a single gene affects the expressions of several other genes or, conversely, the expression of a gene is determined by the effects of many other genes. Thus, it is not sufficient to investigate the expressions of individual genes without analyzing the expressions of other members of a GN. Additionally, the expressions of genes are also influenced by other factors, including the environment, the internal milieu, epigenetic mechanisms, etc. Recent technical advances allow examinations of the expression profiles of a huge number of genes at the same time. Structural genomics (large scale-DNA sequencing), functional genomics (transcriptomics), proteomics (protein chip arraying), and bioinformatics make it potentially possible to map the mechanisms underlying genetic and physiological processes and to explore the machinery thereby these molecular events give rise to phenotype. However, large-scale gene expression profiling techniques have some limitations. First of all, raw expression data reveal only the co-occurrence of particular gene expression patterns, but do not provide information on causal relationships and regulatory hierarchies. Secondly, a change in transcription has a multifactorial background for which *cis*-sequence variation is only one of the factors.

2 The genetic basis of phenotype variation

The term "phenotype" is used in two different senses: a particular trait, such as eye color, and a certain variant of a trait, such as blue eye color. In order to make a clear distinction between these two meanings, we use the term "phenotype variant" for the latter case. While a gene obviously does not encode a phenotype, the one allele, one phenotype variant correlation is not necessarily incorrect, which does not mean that the allele in question encodes the given

phenotype variant. Instead, this relationship merely indicates their invariant co-occurrence. A specific allele—phenotype variant correlation can be a consequence of a lack of recognizable polymorphism in other members of a GN, or the masking of genetic variance due to the robustness of a GN, or it may result from the contribution of the particular gene to the phenotype significantly exceeding the contributions of the rest of the GN components. The question to be considered is the proportion of the phenotypes inherited in this Mendelian-like fashion to those that do not fit this scheme.

2.1 The gene function paradigm—Coding variance in physiological genes

It was earlier believed that the major, if not the exclusive source of variability lies in the sequences of genes encoding physiological proteins, including enzymes, structural proteins, receptors, ion channels, etc. This concept was fostered by the facts that genes and proteins are functional units, and therefore must have undergone positive selection events; and that mutation of genes often results in characteristic phenotypes. The questions to be answered are whether gene functions are still evolving today, or whether this process occurred in the early phase of life's history and was arrested a long time ago. It has been shown that a high degree of polymorphism exists in various enzymes, which are called allelozymes. For example, more than 50 genetic variants of human glucose-6-phosphate dehydrogenase have already been described. However, the vast majority of allelic variants for most genes in a population do not represent functional polymorphism either. Defective genes are exceptions, but they do not contribute to the adaptive variation of a population; in fact, they are subject to purifying selection. Nonetheless, it is not always easy to distinguish dysfunctional alleles from normal gene variants. As another example, Schmid and Tautz[12] compared the sequences of several protein sequences of three *Drosophila* species and found that the fly genomes harbor substantial proportions of genes with a very high divergence rate (one-third of the examined genes were polymorphic), which led to the conclusion that this kind of variation could be a major reservoir for the generation of evolutionary novelties[13]. Current sequence data however, do not support the existence of evolutionary processes that would result in gradual improvements in gene function (including the more efficient catalytic function or improved substrate specificity of a protein) which took place by means of the classic spontaneous mutation/selection scheme. Nevertheless, there are some rare examples that lend support to the role of gradual adaptive changes in some kinds of phenotypic variations. For instance, the affinity of hemoglobin to oxygen in goose species living in high mountains is higher than that of geese living at lower altitudes, which indicates adaptation to poorer oxygen content[14]. Another example is the additive effects of certain amino acid changes at critical sites of opsins, which explain the red-green color vision in certain species[15]. Further, it has been shown that mutations in the melanocortin-1 receptor gene are associated with differential pigmentation in a wide range of species[16]. However, these examples represent interspecific differences. Other exceptions are the paralogous genes; following gene duplication, these may find a novel "physiological niche" that they can fill by co-opting new functions, although changing of regulation of newly duplicated genes is much more common than their functional alterations[17]. Other exceptions are genes that have acquired novel domains by exon duplication or exon shuffling. However, these processes are only of macroevolutionary significance and hardly ever, if at all contribute to the intraspecific variance within species.

2.2 The gene regulation paradigm

2.2.1 *cis*-regulatory polymorphism

The developmental programs of organisms as different as fruit fly and human share a common set of genes. What varies is the amount and the spatio-temporal patterns of gene expression during development. Although, only sporadic data are available at present, it appears that a large-scale functional polymorphism in gene regulation exists among individuals within species. For example, a recent analysis (see Ref. [18]), which compared the upstream regions of several genes of 4 inbred mouse strains, indicated extensive variance in the level and tissue-specificity of gene expression. Similarly, by analyzing the regulatory segments of several genes of a fish species, a substantial polymorphism has been found (see Ref. [19]) in both the regulatory sequences and gene expression levels among individuals. It has also been described that *cis*-regulatory variation in MHC class II promoters in both human and mouse is clearly more frequent than polymorphism in the coding regions of these genes[19]. Extensive polymorphism has likewise been detected in the promoters of genes associated with im-

mune defense[18]. Studies (see Ref. [21]) examining the regulatory sequences of 313 human genes of 82 unrelated individuals, detected a high variability in them. In another survey of regulatory polymorphism, Rockman and Wray likewise reported a widespread functional *cis*-regulatory variance in the human genome[22]. They identified 140 experimentally validated *cis*-regulatory polymorphisms, resulting in a two-fold or greater variation in transcription rate and subsequent gene expression. Other studies[23] analyzed the diversity of proximal promoter polymorphisms on a genome-wide scale of fruit fly genes and showed that they did not differ between strains with divergent gene expression. They concluded that the *trans* sequences might cause differential gene expression. The key question is whether these *trans* effects are caused by structural differences in regulatory genes or by variation in the expression of other GN components, including regulatory genes. At the moment, we practically, have no data on the *cis*-regulatory polymorphism of transcription factors and regulatory RNAs within species.

2.2.2 *trans*-regulatory polymorphism—Coding variance in regulatory genes

(1) Coding variance in transcription factor genes. It has been recently conducted a comparative study[24] on 142 domestic dogs from 92 different breeds, and looked at 37 different tandem repeats in 17 genes in each. The selected genes were developmental transcription factors that play a role in the formation of specific morphologies. Fifteen of the 17 genes proved to have multiple alleles varying in the length of their repeats. A closer investigation of the Runx-2 gene revealed that it is highly polymorphic among breeds (it encodes 18—20 glutamines and 12—17 alanines). The Gln-to-Ala ratio in the alleles of this gene correlated with two craniofacial parameters: the length of the midface and the angle at which the nose bends. Three possible mechanisms have been proposed in an attempt to explain the impact of repeat length on transcriptional activity[25]. Following earlier reports hinting that strings of glutamines drive transcription while polyalanines repress it, it can be hypothesized that high polyGln/polyAla ratios imply strong transcription. Another possibility is that pathological polyAla expansions could induce mislocalization and aggregation, which cannot be efficiently resolved by chaperones. Finally, shorter polyAla repeats may be associated with higher transcriptional activity due to the higher efficiency of the chaperones than that of those of longer repeats to convert these conformation variants to functional protein. An interesting correlation has been observed[24] between the 51 bp deletion in the repeat region of the *Alx-4* gene and the presence of an extra digit in the Great Pyreneese breed. They examined a five-toed Great Pyreneese and found that it had a full complement of bases in the *Alx-4* gene, which was considered verification of their hypothesis. However, dog is a special case since dog breeds display sharply differing phenotypes due to being artificially selected by man. Further investigations are needed to establish whether tandem repeat polymorphism in other genes is also correlated with phenotype variation in dogs, and whether this kind of polymorphism is common and behaves in a similar manner in other species. Interestingly, tandem repeat expansion and contraction occurs at 100000 times higher frequency than point mutations. In addition, due to the involvement of transcription factors, this process can in principle, result in a sudden emergence of specific morphological forms during speciation.

(2) Variance in regulatory RNAs. Recent investigations showed that the majority of mammalian genome is transcribed, commonly from both DNA strands[26,27]. Antisense regulatory RNAs are emerging as the key players in cellular regulation, taking on active roles in multiple regulatory layers. Until now over 1000 micro (mi) RNAs[28] has been described. MicroRNAs can recognize more than one target genes, and a particular gene can be regulated by several miRNAs, therefore, a large number of the genes appear to be under intricate control by these noncoding messages. The type of regulation (degradation through RNA interference pathway or translational repression) by these noncoding RNAs is dependent on the extent of homology between the mRNA and the complementary miRNA. The structure of miRNAs was found highly conserved between species[28]. However, miRNAs are mainly discovered by computational means on the basis of sequence homology (conservation), thus, it is possible that a much larger number of miRNAs exist with lower level of sequence conservation. Another surprising discovery is that a large portion of mammalian genes are regulated by overlapping (cis-) antisense RNAs[29]. Due to their complementarity with the sense DNA strand (encoding mRNA) their structural variations is dependent on the sense strand variation. However, the splicing pattern and the length of overlap between mRNAs

and *cis*-antisense RNAs might vary within species.

Although very little is known of the variability in both the structure and the expression of these noncoding messages, it may be envisaged that accumulating data will lead to a fundamental revision of our perception of the global gene regulatory circuitry of an organism.

Together, physiological proteins do not exhibit significant functional variation within populations. Regulatory proteins (and possibly non-coding RNAs) appear to be exceptions; they may play an important role in the generation of morphological polymorphism within species. This concept seemingly contradicts with the popular view, which holds that sex is originated and maintained to produce genetic variation as a defence mechanism against parasites. It is possible that, although proteins are conserved in their function, specific receptor proteins may vary in their responses to parasites (e.g. certain receptor configurations are not suitable for virus entry to cells).

2.3 The gene network paradigm

The modular arrangement of living systems is typical at all levels of biological organization. In a general sense, a "module" can be defined as a certain type of dynamic pattern of couplings among the constituents of a process, which exhibit a high density of internal interactions and sparse connection to the rest. The autonomous genetic modules are termed variously: gene expression networks, regulatory networks, gene nets, gene networks, genetic pathways, transcriptional regulatory circuits, etc., see Refs. [30—35]. Gene networks can be considered the key modules since they control the formation and operation of anatomical, biochemical, cellular and physiological systems, they are passed on from generation to generation, and they evolve. Furthermore, they are the most basic modules since they cannot dissect further without losing modularity. Gene networks are dynamic systems: their compositions undergo continuous reorganization in response to the shuffling effect of sexual reproduction and to the nonstop alterations in their overall expression profiles as a result of both the interplay between the elements and the influence of the internal milieu and environment[36]. A GN can be regarded as a complex dynamic system that possesses self-organizing features since it contains control mechanisms (positive and negative feedback loops). These mechanisms can confer a robust phenotype[37] on the system, allowing it to buffer perturbations (developmental noise or mutations) and maintain a specific range of activity, often admitting changes in its components without altering its functionalities; alternatively, in specific circumstances this system is capable of major transitions from one state to another one, even in a single step. A GN possesses a hierarchical structure in which the component genes exhibit varying degrees of interconnectivity thereby conferring on the GN a scale-free topology[38], which implies the existence of "Hubs" (the most interconnected genes) that might regulate and integrate the global expression of the GN. GNs are not isolated entities; they crosstalk with and receive feedback from other genetic pathways. Gene networks operate as modules, displaying higher-order collective behavior of the interacting genes. The polymorphism generated by the GNs is dynamic in two senses: the genetic composition of a particular GN in a population undergoes continuous alteration due to the mixing of genetic material by sexual reproduction and in response to evolutionary forces; further, genes exert their effects through a series of dynamically regulated steps via a mutually interdependent interplay between the GN components. The expression profile of a particular gene is multifactorial, involving the architecture of its own *cis*-regulatory sequences (CRSs), the structure, amount and spatiotemporal distribution of *trans*-regulatory elements (transcription factors and regulatory RNAs) directly controlling gene expression, and the expression profiles of other components of GNs exerting their effects through multi-step feedback mechanisms. Additionally, gene expression is dependent on the interplay with other GNs and the environment.

2.4 Epigenetic variation

Phenotype variation can also be produced without a polymorphism in the nucleotide sequence of DNA. Alteration of epigenetic information is manifested in chromatin remodelling, which affects gene expression. The polymorphism of epigenome can be generated by both inherited and environmental factors. Genetic imprinting and maternal effect are means thereby information is transmitted from parents to offspring, while epigenetic alterations during embryogenesis leading to tissue differentiation are genetically programmed cascades of events within the body of an individual. In a recent article Fraga and colleagues[39] examined the epigenetic differences between a large cohort of identical twins. The authors found that, although the twins were epigenetically in-

distinguishable during the early period of life, at older age twins exhibited significant differences in their DNA methylation and histone acetylation pattern. These findings signify the importance of epigenetic processes in the generation of phenotype variation.

3 Afterword

Gene and protein structures are fairly stable even across large evolutionary distances. In contrast, gene regulation appears to vary to a great extent even within species due to the polymorphism in the *cis*-regulatory regions of genes in a gene network. Phenotype variance may also be resulted from the polymorphism in the sequence of transcription factors generated by the expansion and contraction of repetitive trinucleotide sequences. Future data on regulatory RNAs may fundamentally change our views on the mechanism of genetic regulation, and thus the contribution of *trans*- versus *cis*-regulatory variance to phenotypic polymorphism. Besides the above mechanisms, phenotypic differences among distantly-related species are caused by exon duplication, exon shuffling and gene duplication/divergence. The technology for investigation of the simultaneous expressions of a large number of genes is now available, allowing studies of the issue of how given gene expression profiles produced by functionally connected genes give rise to a particular phenotype. At the moment however, our ability to interpret and merge the datasets lags behind our ability to collect them, the reason for this being that expression profiling methods merely reveal co-expression patterns, but tell us nothing about the hierarchical and causal relationships. Traditional molecular biological analyses must therefore be applied to establish causal connections between genotype and phenotype, and hence between genetic and phenotypic variance. Further, computational models must be developed that integrate gene expression data from microarrays with genomic sequence information to explore gene networks and their variability.

References

1 Kimura M. Evolutionary rate at the molecular level. Nature, 1968, 217: 624—626
2 Cowles CR, Hirschhorn JN, Altshuler D, et al. Detection of regulatory variation in mouse genes. Nature Genetics, 2002, 32: 432—437
3 Rockman MV and Wray AG. Abundant raw material for cis-regulatory evolution in humans. Molecular Biology and Evolution, 2002, 19: 1991—2004
4 Fondon JW and Garner HR. Molecular origins of rapid and continuous morphological evolution. Proc Natl Acad Sci USA, 2004, 101: 8053—8058
5 O' Dushlaine CT, Edwards RJ, Park SD, et al. Tandem repeat copy- number variation in protein-coding regions of human genes. Genome Biology, 2005, 6: R69
6 Bucan M and Abel T. The mouse: genetics meets behaviour. Nature Reviews-Genetics, 2002, 3: 414—423
7 Gerlai R. Gene-targeting studies of mammalian behavior: is it the mutation or the background genotype? Trends in Neurosciences, 1996, 19: 177—181
8 Steiner H, Bonner TI, Zimmer AM, et al. Altered gene expression in striatal projection neurons in CB1 cannabinoid receptor knockout mice. Proc Natl Acad Sci USA, 1999, 96: 5786—5790
9 Schumacher J, Jamra RA, Becker T, et al. Evidence for a relationship between genetic variants at the brain-derived neurotrophic factor (BDNF) locus and major depression. Biological Psychiatry, 2005, 58: 307—314
10 Enard W, Przeworski M, Fisher SE, et al. Molecular evolution of FOXP2, a gene involved in speech and language. Nature, 2002, 418: 869—872
11 Boldogkoi Z, Schutz B, Sallach J, et al. $P2X_3$ receptor expression at early stage of mouse embryogenesis. Mechanism of Development, 2002, 118: 255—260
12 Schmid KJ and Tautz D. A screen for fast evolving genes from Drosophila. Proc Natl Acad Sci USA, 1997, 94: 9746—9750
13 Tautz D. 2000 Evolution of transcriptional regulation. Current Opinion in Genetics and Development, 2000, 10: 575—579
14 Jessen T, Weber H, Fermi E, et al. Adaptation of bird hemoglobins to high altitudes: demonstration of molecular mechanism by protein engineering. Proc Natl Acad Sci USA, 1991, 15: 6519—6522
15 Yokoyama S and Radlwimmer FB. The molecular genetics and evolution of red and green color vision in vertebrates. Genetics, 2001, 158: 1697—1710
16 Majerus ME and Mundy NI. Mammalian melanism: natural selection in black and white. Trends in Genetics, 2003, 19: 585—588
17 Carrol SB. Evolution at two levels: On genes and form. Plos Biology, 2005, 7: 1159—1166
18 Mitchison N, Muller B and Segal RM. Natural variation in immune responsiveness, with special reference to immunodeficiency and promoter polymorphism in class II MHC genes. Human Immunology, 2000, 61: 177—181
19 Oleksiak MF, Churchill GA and Crawford DL. Variation in gene expression within and among natural populations. Nature Genetics, 2002, 32: 261—266
20 Cowell LG, Kepler TB, Janitz M, et al. The distribution of variation in regulatory gene segments, as present in MHC class II promoters. Genome Research, 1998, 8: 1—134
21 Stephens JC, Schneider JA and Tanguay DA. Haplotype variation and linkage disequilibrium in 313 human genes. Science, 2001, 293: 489—493
22 Rockman MV and Wray GA. Abundant raw material for *cis*-regulatory evolution in humans. Mol Biol Evol, 2002, 19(11): 1991—2004
23 Brown RP and Feder ME. Reverse transcriptional profiling: non-correspondence of transcript level variation and proximal promoter polymorphism. BMC Genomics, 2005, 6: 110
24 Fondon JW and Garner HR. Molecular origins of rapid and continuous morphological evolution. Proc Natl Acad Sci USA, 2004, 101 (52): 18058—18063
25 Caburet S, Cocquet J, Vaiman D, et al. Coding repeats and evolutionary "agility". Bioessays, 2005, 27, 581—587

26 Caminci P, Kasukawa T, Katayama S, et al. The transcriptional landscape of the mammalian genome. Science, 2005, 309: 1559—1563
27 Mattick JS and Makunin IV. Non-coding RNA. Human Molecular Genetics, 2006, 15: R17
28 Pasquinelli AE, Hunter S and Bracht J. MicroRNAs: a developing story. Current Opinion in Genetics and Development, 2005, 15: 200—205
29 Chen J, Sun M, Kent WJ, et al. Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Research, 2004, 32: 4812—4820
30 Bonner JT. The Evolution of Complexity. New Jursey: Princeton University Press, 1988
31 Raff R. The shape of life. In: Genes, Development and the Evolution of Animal Form. Chicago: University of Chicago Press, 1996
32 Cantile M, Procino A D'armentino M, Cindolo L, et al. HOX gene network is involved in the transcriptional regulation of *in vivo* human adipogenesis. Journal of Cellular Physiology, 2003, 194: 225—236
33 Davidson EH, Rast JP and Olivery P. A genomic regulatory network for development. Science, 2002, 295: 1669—1678
34 Scannell DR and Wolfe K. Rewiring the transcriptional regulatory circuits of cells. Genome Biology, 2004, 5: 206—210
35 Wagner GP. Homologues, natural kinds and the evolution of modularity. American Zoologist, 1996, 36: 36—43
36 Boldogkoi Z. Gene network polymorphism is the raw material of natural selection—The selfish gene network hypothesis. Journal of Molecular Evolution, 2004, 59: 339—356
37 Kitano H. Biological robustness. Nature Reviews—Genetics, 2004, 5: 826—837
38 Barabasi AL and Bonabeau E. Scale-free networks. Scientific American, 2003, 288: 60—69
39 Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci USA, 2005, 102(30): 10604—10609