ISSN: 2279 - 0594

Available Online at www.jbpr.in

Journal of Biomedical and Pharmaceutical Research 3 (1) 2014, 07-13

REVIEW ARTICLE

PROMINENCE OF HUMAN PHENOME IN HUMAN GENOME

Pramod Singh Khatri[1], General Mahavir Singh[2], Sumit kumar[3]

[1] HOD, Clinical Research, Amity Medical School, Amity University, Haryana, India.

[2] HOI, Amity Medical School, Amity University, Haryana, India.

[3] Associate Professor, IAMR, Ghaziabad, India.

Received 12 December 2013; Accepted 20 December 2013

ABSTRACT

The phenome is a whole set of phenotypes ensuing because of hereditary variety in populaces of an organism. Immersion of a phenome intimates the distinguishing proof and phenotypic depiction of changes in all genes in a creature, conceivably obliged to those encoding proteins. The human genome is thought to hold 20-25,000 protein coding genes, yet just a little portion of these have recorded mutant phenotypes; therefore the human phenome is a long way from complete. This paper analyse the common state of the human phenome, and talks over hypothetical and viable contemplations to an immersion dissection in people. All around, stress is set on high penetrance hereditary variety, of the kind ordinarily co-partnered with monogenic versus complex characteristics.

Keywords: Genetics, Mutation, Genome, phenome.

INTRODUCTION:

The idea of phenomic immersion can't be approached without a watchful attention of the significance of phenotype, and the genotype-phenotype outline. In common, the phenotype of a singular organic entity is the entirety of its physiology. In heredity framework, phenotypes may be characterized all the more particularly as those physiological qualities which change measurably as a capacity of genomic arrangement contrasts around people in a populace. Thus, numerous angles of physiology can be overlooked depending on the sort of succession distinctions being contemplated. Case in point, transformations in the gene encoding the low density lipoprotein receptor (LDLR) have gigantic effect on plasma cholesterol, atherosclerosis, coronary illness, yet have no clear or known effect on eye color or tallness. Conversely, changes in the genes encoding pigment transforming catalysts in the iris may be unimportant to coronary illness (Paulson et al, 2009). One of the points of interest of forward heredity is that one starts with an unexpected physiological characteristic watched in human life, and continues to look for hereditary variety that underlies that attribute. conversely, if one begins with a interested gene , and transforms it in a controlled manner, one has no clue what the physiological impacts could be so one is gratified to investigate all conceivable phenotype outcomes – a period devouring and troublesome assignment ,yet human heredity in whole organic entities is by and large confined to forward heredity with the exception of in some uncommon instances of clinical trials including substantial cell hereditary control (i.e. gene therapy).

Monogenic Contrasted with Multifaceted Phenotypes:

To some degree, the difference between monogenic and multifaceted phenotypes is effective. All hereditary variants in chromosomes will be de facto transferred in a Mendelian style. Hence even pitifully penetrant variants, for example, most SNPs found by companionship through vast scale genome- wide examines in the event that control cohorts, are in any case "Mendelian". Monogenic or Mendelian conditions are characterized as being initiated by deviations in single genes, with the variations having a high penetrance (i.e. included risk). As a basic illustration, the scientific phenotype of cystic fibrosis has been made for numerous patients over all human populaces (fig1). In all cases that have been contemplated, changes are found in a solitary protein-coding gene, CFTR. Therefore there is an agreeable one gene-one phenotype connection in cystic fibrosis condition (Dror et al 2009). Conversely, oligogenic or polygenic or complex phenotypes are typically considered those for which consolidated input of numerous distinctive hereditary variations is needed,

*Corresponding author: Pramod Singh Khatri| Email: pramodsinghkhatri@gmail.com

conceivably with critical ecological components needed as well.

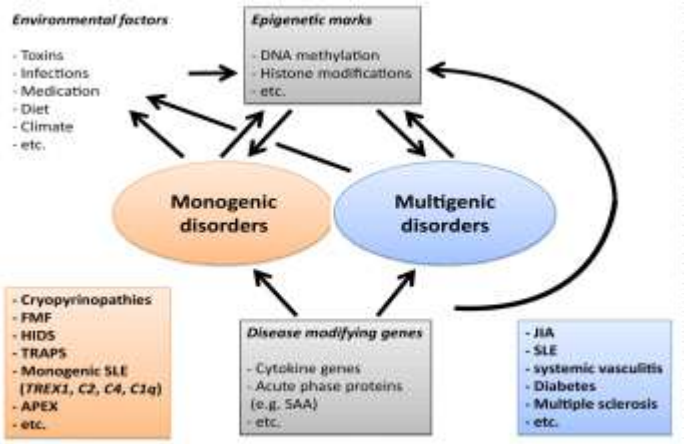

Figure 1: Monogenic vs. Multigenic disorder

Actually for ostensibly monogenic syndrome the instance of CF is far from ordinary. There will be numerous restorative instances in which a specific clinical finding has been joined to high penetrance hereditary variety in numerous distinctive genes (e.g. ataxias, sensorineural deafness, retinitis pigmentosa). In Charcot-Marie-Tooth sickness, contributory genes have been found with an expansive range of cell forms incorporating specific neuronal structures, protein turnover, vesicle combination, microtubule transport, transcription variables, and even a few tRNA-synthetases. Obviously, some syndrome judgments represent to a normal close focus for distinctive sorts of imperfections in gene function.

There will be additionally a mixed bag of moderately decently recorded illustrations of digenic inheritance in human sickness, such that a clinical phenotype relies upon synchronous hereditary variety in two distinctive protein-coding genes. Retinitis pigmentosa coming about because of concurrent heterozygous changes in the genes peripherin/RDS and ROM1 is the initially recorded instance. There are likewise likely digenic models for hemochromatosis, Bardet-Biedl syndrome, sensorineural deafness and different phenotypes(Freimer et al ,2003).

Monogenic syndromes are operationally distinguished principally by their watched transmission in families. On account of feebly penetrant hereditary variants, or those with major hereditary modifiers, it is challenging or unthinkable to follow the inheritance of the physiological state (if a sickness, a quantitative characteristic, or a benevolent attribute, for example, stature) in families, in a manner dependable with a solitary chromosomal locus. Besides, for each narrative instance of qualities "running in relatives, for example, hair or eye color or tallness, there are instances where this does not happen. Plainly, some physiological states require different hereditary inputs of comparative quality penetrance, and some allelic variants on their own have minimal physiological sway. Accepting that all physiological qualities can be characterized by some quantitative variables, in spite of the fact that these may not be known, the impact of a singular variant allele could be pictured by its impact on the mean quality estimation(Reilgert et al ,2009). The further from the populace mean a specific allelic variant is, the less demanding it is to recognize and follow in unique families. The qualification between prevailing (dominant) and latent acting (recessive) alleles, while a commonsense inconvenience in the identification of familial transmission, is hypothetically not pertinent to this fact.

Not all attributes are ailments, clearly, despite the fact that maladies get the most consideration from geneticists and are regularly the most promptly determined through routine clinical reconnaissance. In any case, non-medical characteristics, for example, hair or eye color and capability to smell or taste specific chemicals, have been examined hereditarily.

Total Number of Genes in Genome

Figuring out the amount of real genes encoded by human (or other) genomes has demonstrated shockingly frustrating. From early evaluates in overabundance of 80,000 protein coding genes in the human genome dependent upon quantitative estimations of CpG islands, gauges have dropped over and again, to 28-32,000 built with respect to exon examination between the pufferfish and human genomes, to 21,037 dependent upon vast scale cDNA library sequencing , to 20,488 built in light of correlation of proposed human genes to sequences of other primate genomes. As of now RefSeq, a broadly utilized NCBI library of gene annotations incorporates 21,515 interesting archives. These incorporate generous amounts of non-protein coding genes, for example, 363 small nucleolar RNA (SNO) genes, 28 SNAR genes and 637 micro- RNA (MIR) genes, so RefSeq presumably incorporates roughly 20,000 protein coding genes. There are numerous progressing manual gene curation deliberations, particularly Havana, VEGA and CCDS and also the Mammalian Gene Collection (MGC). The CCDS has presently curated 18,173 diverse protein coding genes, despite the fact that some decently reported genes will be obviously missing from the database which is therefore still in advancement. The MGC as of now incorporates 17,592 genes for which full length cDNA clones are accessible (oti et al,2009).

Despite the fact that a consensus shows up close to that the sum number of protein coding genes in the human genome is between 20-22,000, the approach of next generation innovations has permitted further investigation of elective grafting through profound resequencing of cDNA libraries. Some latest studies have reported novel alternative joined types of known genes, and totally novel exons inside formerly contemplated genes. Therefore the amount of potential elective protein isoforms is numerous times more excellent than the real gene tally.

The human mitochondrial genome has been relatively vibrant to clarify, as it looks like bacterial genomes in gene structure. It encodes 13 proteins, all included in respiratory electron transport and oxidative phosphorylation, as well as 22 mitochondrion particular tRNAs, and the 12S and 16S ribosomal RNAs (rRNAs) needed for the mitochondrial ribosome . In any case, the human mitochondrion itself holds 900-1000 diverse proteins, in light of immediate proteomic breakdowns. Therefore the expansive lion's shares of mitochondrial proteins are encoded by chromosomal genes.

Till date total monogenic disorder known?

Remembering these focuses, what numbers of described monogenic phenotypes are there in population? The whole human heredity prose is boundless. There will be unique locus particular databases or sites for numerous hereditary syndromes, for example, cystic fibrosis and familial hypercholesterolemia. There are three associations endeavoring to clergyman the whole literary works. The best known, Online Inheritance in Man (OMIM), function as a blended therapeutic and hereditary database, with sections for clinical syndrome mixed with entries for particular genes . It has restricted immediate seek functions. Questioning with the widespread term 0001 together with an utmost for allelic variations, OMIM returns 2420 free gene sections at the time of composing, while the facts page reports 2763 phenotypes with a known molecular origin. Practically the sum of these signifies monogenic syndrome (high penetrance variations with sensible proof of genic pathogenicity dependent upon the transformation itself). Though, a part of these, perhaps the same amount as 10%, may represent to low penetrance chance alleles not plainly quantifiable as a feature of monogenic syndrome (Berger et al,2010). OMIM does not minister all described hereditary variations in all genes, in spite of the fact that entries are customarily overhauled when experimentally intriguing results are published.

The Human Gene Mutation Database (HGMD) developed at Cardiff University, endeavors to unite all reported hereditary variations in literature. It has a fragmented freely accessible form, and an updated business rendition. The business HGMD right now incorporates 3611 autonomous gene sections and 96,631 transformation entries. It is challenging to survey what number of these relates to monogenic conditions, versus low penetrance risk aspects for multifaceted phenotypes. HGMD recommends that 2491 genes hold at least one possible high penetrance allele (frameshift or premature stop codon, splice site, missense or indel. The distinction of less than 100 genes between OMIM and HGMD may identify to intricacies of written works curation of recommended versus affirmed genotype/phenotype causation.

The Human Genome Variation Society (HGVS) is preparing a project of exhaustive curation which might inevitably give comparative data as OMIM and HGMD joined together. Presently the database HGVbaseG2P seems centered essential on the effects of GWAS studies for complex syndrome. It incorporates connections to HGMD, yet apparently just the openly accessible content not the full restrictive database.

The sum number of molecularly described, correct monogenic syndrome remains somewhat questionable, however is marginally less than 2500, signifying to 10-12% of protein- coding genes. For lingering ~90% of genes, no high penetrance hereditary variety has been causally connected to a physiological phenotype in people.

Phenome Mutation Rates:

Proportions of spontaneous transformation in the human genome have been assessed in different ways. Merging early consequences of sequencing a little number of genuine or pseudogenes. By analyzing fractional human and chimpanzee arrangements, it was estimated that impartial rate of $1.3\text{-}3.4 \times 10^{-8}$ events for every bp for generation, equal to 91-238 new transformations for every zygote. Recently a direct test was performed by sequencing substantial sections of flow sorted Y chromosomes from two guys of an exactly known lineal gene a log y. Four new changes were located between the two specimens, which amending for chromosome size yielded a rate of 3×10^{-8} events for every bp for generation, in spite of the fact that the little number of unique occasions suggests a sensible sample predisposition . Finally, entire genomes were resequenced for a family of two parents and two posterity, permitting the most immediate conceivable measure. Total 28 affirmed new transformations were

found, which after different rectifications yielded a rate of $1.1x10^{-8}$ events for every bp for generation, proportional to 70 new events for every zygote, with a generously lower risk of specimen inclination

These all indicate point transformation rates. It is presently clear that bigger structural rearrangement known as (CNV), likewise emerge at sensibly high frequencies in human populaces, in spite of the fact that it is challenging to evaluate accurate de novo transformation rates(Phelen et al ,2000). Many CNVs incorporate genes and have potential function.

Not all new transformations need have functional significances. It is recognized that the majority of them are indeed practically impartial. There is no simple approach to evaluate this since most new transformations emerge in non-coding regions of the genome, in either introns or intergenic regions, where our comprehension of functional components is still exceptionally inadequate. The part of haploid sequence coding for protein is normally acknowledged to be 1-2%, subsequently if every zygote gets 70 new transformations, at most one of these is required to be in a coding exon. Coding variations may be silent (synonymous) or change an amino acid (nonsynonymous). Some studies have contended that extent of nonsynonymous changes have functional implications, despite the fact that these might just be apparent over evolution time frames and not be straightforwardly pertinent to elucidation in people . It ought to be noted that synonymous changes might likewise have measurable impacts on gene function, by impacting splicing through exon splice enhancer or repressor elements, or translation effectiveness

Dominance Contrasted With Recessive, Gain As Opposed To Loss of Function:

The extent of monogenic syndrome transmitted as predominant versus latent (recessive) qualities in families is tricky to figure out from the curated databases (fig2). Thus, it is troublesome to evaluate the extent of gain of function (gof) versus loss of function (lof) alleles(fig3), despite the fact that this is not absolutely an issue of curation since for numerous alleles the utilitarian impact remains unconfirmed biochemically. Complete gene erasures (deletions) speak to the gold standard for a (lof) allele, however few clean deletion of distinct genes have been connected with monogenic conditions (although numerous adjacent deletion syndromes enveloping multiple genes are generally recorded). Geneticists usually expect that most important protein truncations emerging through frameshift or untimely stop codon changes(mutation) will be pathogenic, yet it might be credulous to assume that these all speak to complete (lof) since amino terminal protein pieces could effectively encode halfway or unusual activity. For distinct genes, the most ideal approach to recognize between addition and misfortune of function phenotypes is to perceive the array of alleles partnered with a specific phenotype. for instance, transformations of BRCA1 which expand the danger of early onset bosom tumor incorporate a wide mixed bag of missense, truncating and deletion transformations

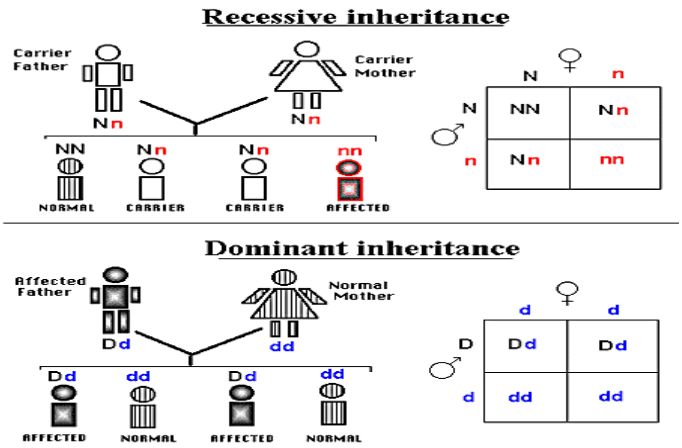

Figure 2: Dominant inheritance vs Recessive Inheritance

It will be sensible to assume that most of these signify (lof) alleles. Similarly, numerous free alleles bringing about Huntington syndrome all include extension of the triplet rehash, while causal changes somewhere else in the gene are not known for this phenotype, steady of the elucidation that these are gain of function(gof) alleles. In contrast, Friedreich ataxia will be commonly twisted by a triplet rehash extension in the frataxin gene, however a little division of patients come about because of point alleles, both missense and truncating. Possibly, the ataxia epitomize to a loss of function (lof) phenotype, and the recurrence of triplet rehash expansion alleles could come about basically from an effortlessly accessed mutational mechanism versus a particular novel function of the transformed protein. On the other hand, some particular point alleles could cause an abnormal novel protein function likened to that of the triplet rehash expansion. One glaring case is the gene encoding superoxide dismutase (SOD), transformations in which are causal for a part of instances of amyotrophic lateral sclerosis (ALS, or Lou Gehrig Syndrome).

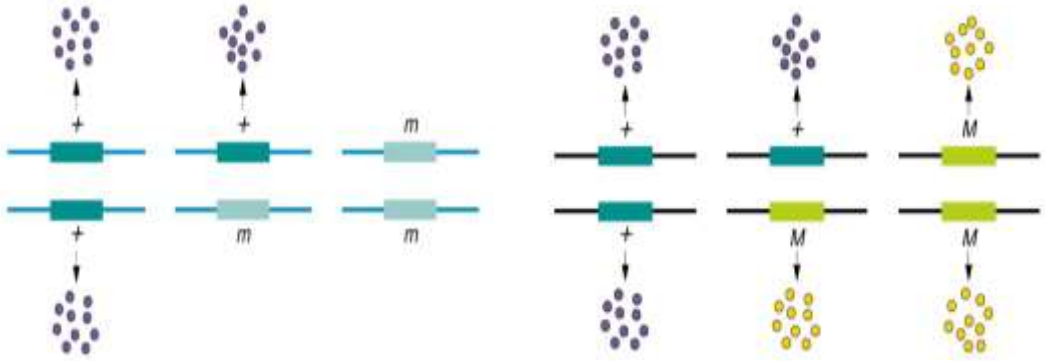

Figure 3: (a)Loss of function (lof) mutation, (b) Gain of Function(gof) mutation.

There are numerous distinctive causal changes in SOD, practically all of which will be missense variations. In any case, a rodent gene knockout does not create the ALS phenotype, inasmuch as overexpressing identical missense variations summarizes some viewpoints of the human infection. Therefore it can't be expected that phenotypes emerging from numerous diverse transformations in a gene consistently express to misfortune of function, in spite of the fact that they likely do by and in large situations.

The latent(recessive) nature of most changes was acknowledged early in the twentieth century, and hypothetical geneticists prepared numerous elective demonstrations to illustrate this . Expecting that a biochemically quantitative pathways cause all physiological traits, even those of embryonic improvement or neuropsychiatric capability, metabolic control hypothesis recommends anyhow in principle that gene measurement reaction could promptly be nonlinear, and that loss of even a full duplicate of gene function may have just incomplete impacts in non-enzymatic frameworks (Scriver et al,2004).

Outlining Lingering Phenome:

If a little part of protein-coding genes has a molecularly portrayed monogenic phenotype in people as of now, what will this mean for the whole phenome? Assuming that the complete phenome contains all conceivable phenotypes of all genes in the genome, and then the phenome is successfully boundless. Be that as it may, if most loss of function (lof) alleles in a assumed gene cause comparative phenotypes, then the lof phenome is much more modest, presumably limited, and just somewhat bigger than the amount of genes. Further, if gof alleles are much rarer than lof alleles, then the greater part of the gof phenome will never be watched given the aggregate human populace size. Interestingly, as depicted, the lof phenome likely exists over the planet close finishing now.

All these focuses together unequivocally infer the accompanying situation: lof alleles of all genes in the genome emerge customarily around the globe; the greater part of these speak to latent (recessive) alleles; clinical ascertainment of homozygotes for these alleles will be separately extraordinary with the exception of in founder or other populace sorts with higher rates of consanguinity than needed for complete panmixis; compound heterozygotes will emerge at low rates basically all around, with almost no expanded rate in founder populaces. The genuine rate of manifestation of passive (recessive) cases that may be clinically discovered is exceptionally challenging to gauge, since it depends on current allele frequencies for malicious alleles, which thusly depend incompletely on items of neighborhood populace demographic history which for most populaces are crudely known. On the other hand, even in hypothetical outbred or founder populaces there can be expanded recurrence of latent pernicious alleles through neglected low levels of populace versatility, the diminished number of watched versus potential haplotypes found in non-African populaces furnishes a compelling instance of this(Matsunaga et al,2009).

The approach of thick SNP marker boards has significantly encouraged the mapping of latent(recessive) hereditary loci, particularly in populace secludes with not just expanded recessively additionally homozygosity. Formal linkage dissection is moderately wasteful for recessive issue unless there are a curiously expansive number of influenced people in a given family or identified populace (i.e. founder impacts).

Operationally, doling out genotype/phenotype causality explicitly requires either various patients transformed in the same gene with a uncommon monogenic phenotype, or then again solid acceptance of a hopeful

gene through creature model studies. Regardless, entire genome or exomes sequencing is ready to upset medicinal heredity, permitting the recognizable proof of causal genes for numerous more extraordinary phenotypes, incorporating some, (for example, pre-birth lethality, or predominant illness with diminished practicality or fruitfulness) for which family studies are naturally incomprehensible.

Theoretical Solicitations of Phenome Immersion:

Above the modestly investigative value of reckoning out high penetrance mutational phenotypes for as many genes as could be expected under the circumstances, there is likewise reasonable esteem in this program. The human heredity group has centered much of its exertion for as far back as decade on characterizing hazard variants for complex, frequently mature person onset degenerative illnesses of incredible social and business criticalness. This work has generally utilized the explanatory standard of high throughput SNP genotyping and case/control examination utilizing substantial patient accomplices (the GWAS approach). GWAS studies have succeeded in distinguishing numerous such hazard components, however in general the variants found exclusively include just little incremental chance of malady, additionally just an unobtrusive division of aggregate hereditary chance in populaces is logical summing over all such normal alleles . The rising accord is that uncovering the "misplaced heritability" will require high throughput genomic resequencing in bigger and bigger companions. Assuming that one of the objectives is to distinguish novel restorative focuses for pharmaceutical mediation; this will be a long haul and momentously unreasonable proposit.

Conversely, there are numerous cases of high penetrance monogenic attributes of immediate pertinence to regular sickness. These might be found either by universal family-based linkage studies , or for quantitative characteristics, for example, figure mass, plasma glucose or cholesterol, by sequencing populace outliers searching for overabundance variety in particular genes. Basically, some segment of the missing heritability for medically vital attributes lies in an extensive number of independently uncommon alleles (van soest et al,1999). Such genes furnish "low hanging soil grown foods" for drug improvement, something wildly required by the biotechnology and pharmaceutical enterprises, which will be generally recognized to be enduring a shortage of exceptional new items. For instance, the human genome encodes around 370 G-protein coupled receptors, not incorporating olfactory receptors. A considerable lot of these are focuses of medications as of recently available, as GPCRs have a few major preferences for new drug development process. As with the human genome all in all, just in the vicinity of 10% of GPCRs have archived human monogenic phenotypes. Many of these human phenotypes are specifically significant to critical malady models. Thus the GPCR phenome speaks to a fruitful region for hereditary dissection, obliging just the resequencing of this set of genes in clinically characterized instances of therapeutic entomb.

CONCLUSION:

we have endeavored to record that most of the human phenome stays to be discovered and molecularly portrayed. This is basically because of tests in clinical ascertainment & gene mapping, not to any hypothetical issues with inherent malicious change rates or the event of long ago uncharacterized phenotypes. Change rates in the human populace are sufficiently high that pathogenic variation must exist in all practical genes at the present time. The vast majority of these are liable to cause latent (recessive) phenotypes; hence ambiguity depends on distinguishing homozygous or compound heterozygous transformation bearers from around all medically recognized patients. Generally, causal gene distinguishing proof has hinged upon the Identification of complex families, emulated by mapping with boards of unnamed DNA polymorphisms, position cloning and resequencing. With the approach of cutting edge sequencing advances, it is currently plausible to sequence entire genomes or exomes of patients without the need for families or linkage mapping. On the other hand, the huge measure of singular genome variety raises the task of distinguishing causal variations in these patients. This will oblige joint effort around geneticists scattered around the planet, perhaps with advancement of huge genotype/phenotype databases, together with the utilization of model living beings to approve gene Function or transformation pathogenicity.

Acknowledgement:

We wish to thank Prof. (Dr.) Gen. Mahavir Singh for his critical discussion.

Conflicts of Interest Statement:

The Authors declare no conflicts of interest.

REFERENCES:

1. Freimer, N.; Sabatti, C. The human phenome project. *Nat. Genet.,* 2003, *34*, 15-21.
2. Scriver, C.R. After the genome the phenome? *J. Inherit. Metab.Dis.,* 2004, *27*, 305-317.

3. Paulson, H.L. The spinocerebellar ataxias. *J. Neuroophthalmol.,*2009, *29*, 227-237.
4. Hilgert, N.; Smith, R.J.; Van Camp, G. Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics? *Mutat. Res.,* 2009, *681*, 189-196.
5. Oti, M.; Huynen, M.A.; Brunner, H.G. The biological coherence of human phenome databases. *Am. J. Hum. Genet,* 2009, *85*, 801-808.
6. Dror, A.A.; Avraham, K.B. Hearing loss: mechanisms revealed by genetics and cell biology. *Annu. Rev. Genet.,* 2009, *43*, 411-437.
7. Matsunaga, T. Value of genetic testing in the ontological approach for sensorineural hearing loss. *Keio J. Med.,* 2009, *58*, 216-222.
8. Van Soest, S.; Westerveld, A.; de Jong, P.T.; Bleeker-Wagemakers,E.M.; Bergen, A.A. Retinitis pigmentosa: defined from a molecular point of view. *Surv. Ophthalmol.,* 1999, *43*, 321-334.
9. Berger, W.; Kloeckener-Gruissem, B.; Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Eye Res.,* 2010, *29*(5), 335-375.
10. Phelan, J.K.; Bok, D. A brief review of retinitis pigmentosa and the identified retinitis pigmentosa genes. *Mol. Vis.,* 2000, *6*, 116-124.

