The Egyptian Journal of Medical Human Genetics (2015) **16**, 101–116

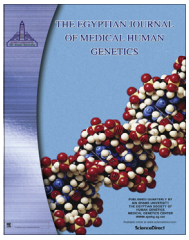
HOSTED BY

ELSEVIER

Ain Shams University

The Egyptian Journal of Medical Human Genetics

www.ejmhg.eg.net
www.sciencedirect.com



REVIEW

Formal genetic maps

Mohammad Saad Zaghloul Salem

Professor of Medical genetics, Faculty of Medicine, Ain-Shams University, Cairo, Egypt

Received 20 November 2014; accepted 25 November 2014
Available online 24 December 2014

KEYWORDS

Physical genetic maps;
Functional genetic maps;
Experimental genetic maps;
Constructed genetic maps;
Transposon maps;
Telomere maps;
Proteome maps;
Oncoprotein maps;
Tumor maps;
Regulatory networks maps;
Genotype/haplotype maps;
Disease association maps

Abstract Formal genetic maps are databases, represented as text or graphic figures, that can be collected/organized/formulated and constructed for nearly any, and every, structural or functional region of the genetic material. Though these maps are basically descriptive, their analysis can provide relevant crucial data that can be applied for different purposes in many fields. The more comprehensive these maps are the more significant information that can be provided through their analysis. Formal genetic maps comprise four main categories: *physical maps* detailing the structural characteristics of different regions/sequences of both the nuclear and the mitochondrial genomes, *functional maps* describing the varied functional potentials of the different components of the genome/transcriptome/proteome, *experimental induced maps* that are intentionally designed and constructed for specific purposes and *constructed maps* that are deduced and extracted from other formal maps to serve particular targets that cannot be achieved solely by the constituent maps. Formal genetic maps have a wide spectrum of applications in all fields of human genetics including basic genetics as well as medical genetics. The beneficial impact of the different types of formal genetic maps imposed their application in nearly all fields of medical genetics including basic/clinical/diagnostic/therapeutic/prophylactic and applied genetics and made these maps indispensable tools in research studies relevant to these fields.

Contents

Peer review under responsibility of Ain Shams University.

http://dx.doi.org/10.1016/j.ejmhg.2014.11.006

1. Definition

Formal genetic maps (Table 1) are databases aiming at providing important, beneficial and crucial clues relevant to nearly all fields of human genetics, including medical genetics. Progress in delineation and analysis of structural organization of the human genome generates a flood of information leading to characterization of new formal maps of specific DNA markers and regions of both structural and functional significance. These maps represent bioinformatics databases that have a crucial impact on many aspects of basic as well as clinical medical genetics.

2. Types of formal genetic maps

Formal genetic maps comprise two main groups: *physical genetic maps* and *constructed genetic maps*. Physical genetic maps reflect data that represent naturally occurring physical situations of the genetic material, like chromosomal locations of genes or chromosomal maps. *Constructed genetic maps*, on the other hand, comprise data observed, collected and analyzed from research studies, e.g., linkage maps, as well as data generated through experimentally induced situations, e.g., radiation hybrid maps. However, it must be emphasized that apart from some physical genetic maps, the differentiation between the two groups of genetic maps is not absolute, considerable overlap between physical and constructed maps do exist and some types of genetic maps can be considered within the context of both groups. For instance, linkage maps are constructed based on data analyzed from inbreeding experiments or offspring genotypes, but they also reflect physical situations of genes showing linkage during recombination. Also, oncogene maps, for instance, can be considered within the context of physical genetic maps since they represent specific structural components of the genome as well as within the context of constructed genetic maps because they represent data collected and constructed for particular purposes, e.g., diagnostic approaches to cancer and predictive/prophylactic approaches to individuals prone to have malignant tumors. Many different types of genetic maps can be arbitrarily designed and constructed for specific targets. *Functional genetic maps*, for example, can be formulated to delineate specific regulatory/intermediary/executive functions of specific structural

Table 1 Formal genetic maps.

Physical maps	Functional maps	Constructed maps	Induced/experimental maps
1. Genome maps	1. Transcriptome maps	1. Linkage maps	1. Inter-species hybridization maps
2. Gene maps	2. Regulatory RNA maps	2. Genotype maps	2. Restriction fragment length polymorphism maps
3. Transposon maps	3. piwiRNA maps	3. Haplotype maps	3. Radiation hybrid maps
4. Pyknon maps	4. Proteome maps	4. Phenotype maps	4. Probe-specific maps
5. Pseudogene maps	5. Oncoprotein maps	5. Disease association maps	
6. Telomere maps			
7. Chromosome maps			

components of the genome/transcriptome/proteome for particular aims. Proteome maps, for instance, can be constructed based on proteome components of distinctive genetic disorders for diagnostic/prognostic/predictive purposes. They can also be constructed by delineating proteome profiles of particular genomes for comparative characterization of different proteomes of studied organisms. *Structural/functional genomic maps* are comprehensive databases regarding critical structural components of the genome that mediate vital constitutive and regulatory functions responsible for maintaining many crucial aspects of the genome including genome identity, integrity and stability. These maps include pyknon maps, pseudogene maps, transposon maps and telomere maps.

3. Classification of formal genetic maps

Formal genetic maps can be classified into four main categories that are widely used in nearly all fields of basic, clinical, diagnostic and experimental genetics. They include: the *family pedigree*, *physical maps*, *functional genetic maps*, *constructed genetic maps* and induced or experimental genetic maps (Table 1). In addition, extension of the use of some of these maps can have a marked impact on other fields of medical genetics like therapeutic, prophylactic and applied genetics. The number of genetic maps constructed based on certain data related to the structural organization and functional specialization of the genetic material is, apparently, unlimited. Any significant genetic data can be mapped and used for specific purposes. Construction of bioinformatic databases and comparative analysis of data represent the main applications of formal genetic maps.

4. Applications of formal genetic maps

The bioinformatic databases are represented as, and included within, different types of formal genetic maps that have a wide range of applications in many fields of basic, clinical, diagnostic, therapeutic, prophylactic and applied genetics. Construction of the *family pedigree* of patients with genetic diseases, or of families seeking counseling advice, constitutes the first step in approaching patients and families having genetic disorders, and represents the simplest and most direct of these applications. Genealogical analysis of data represented by informative symbols of the pedigree allows for deriving relevant genetic information, like the possible pattern of inheritance, and to calculate recurrence risk figures in future offspring.

Progress in analysis of the structural organization of the human genome generates a flood of information leading to characterization of new formal maps of specific DNA markers and regions of both structural and functional significance. These maps represent *bioinformatic databases* that can have a crucial impact on many aspects of basic as well as of clinical medical genetics. For instance, *exome maps* comprising detailed information of exons of genes can be constructed and used for both intra and inter-species comparative purposes. Similarly, within the context of pathogenetics, comparison of exome maps of patients with specific idiopathic genetic disorders with those of normal subjects might represent a promising approach that can have many predictive and diagnostic applications in clinical genetics. Other types of molecular maps that can be constructed based on available as well as on the rapidly accumulating databases of human genome structure, e.g., *introme* maps, *pyknon* maps, *transposon* maps, *telomere* maps and maps of *pseudogenes*, can also have a wide spectrum of applications in many fields of medical genetics.

The significant beneficial effects and applications of *structural genetic maps* in different fields of medical genetics call for construction of parallel databases of *functional genetic maps* that characterize critical functional markers and transcriptionally active regions and sequences of the human genome. Examples of such functional maps can, for instance, include *proteome maps* comprising both structural protein and enzyme, or catalytic, protein databases. Comparative analysis of these functional protein maps in normal subjects and in patients affected with specific genetically-determined disorders and idiopathic diseases caused by, still, unidentified etiological mechanisms, might prove helpful in revealing the underlying *pathoproteomic abnormalities* responsible for the development of the specific pathophysiological alterations that characterize the clinical phenotype of each of these diseases. In addition, in a way similar to that of reverse engineering, comparative analysis of normal proteome maps and abnormal proteome maps of specific genetic disorders can disclose underlying pathogenetic mechanisms, pathotranscriptomic differences and causative genetic mutations possibly involved in mediating the pathogenesis and development of these disorders.

Hypothetical applications of functional human genomic maps in medical genetics can include a wide range of countless diagnostic techniques and therapeutic approaches. Comparative analysis of *proteome maps* of normal cells and those of malignant cells can, for instance, depict functional deviations of the genome in cancer cells and detect oncogenic proteins synthesized and expressed by the malignant cell and responsible for pathogenesis and evolution of the malignant phenotype. *Oncoprotein maps* constructed from this information would, certainly, have beneficial impact on research studies aiming at better, and proper, understanding of malignant transformation through targeting pivotal oncoproteins that initiate and promote the malignant phenotype of cancer cells. They would, also, have similar significant impact on research studies aiming at designing and tailoring specific therapeutic trials for challenging carcinogenesis through e.g., engineering of monoclonal antibodies against relevant oncoproteins inside intracellular compartments, as well as on cell membranes and within the intercellular environment. Interruption of signaling pathways involved in the mediation of many selective advantageous functions of malignant cells, e.g., enhanced glycolysis, metastasis and neovascularization, which depend on the synthesis of specific oncoproteins could, probably, represent a promising approach in treatment of cancer via disassembly of metabolic and regulatory networks actively participating in malignant transformation and evolution of the malignant phenotype.

On the molecular level, comparative analysis of *transcriptome maps* of messenger RNA and of different micro RNA species in normal and in diseased subjects could offer remarkable help in the diagnosis of genetic diseases characterized by substantial genomic activities. Malignant cells express large numbers of actively transcribed genes and detailed cancer-specific transcriptome maps could be constructed for pathologically distinct types of malignancies, thus allowing for better

understanding of the aberrant phenomena of differential suppression of certain genes and of mass overexpression of specific gene families, e.g., proto-oncogenes, in malignant cells. They can also offer more information on the genic and inter-genic pathogenetic and pathological mechanisms underlying the development of specific types of cancer. Cancer-specific transcriptome maps and oncoprotein maps constructed for pathologically distinct types of malignancies might offer rapid and accurate diagnostic tools of these malignancies when compared with corresponding maps of normal cells. Delineation of the whole spectrum of proto-oncogenes and of oncogenes in the human genome and construction of informative maps detailing relevant databases of this spectrum remains, and represents, a prospective ultimate goal crucial for understanding and designing radical therapies for cancer.

5. Developmental formal genetic maps

During embryonic and fetal development, and to much lesser extent during early post-natal life, organ and tissue-specific differential suppression of large numbers of certain sets of genes, paralleled by selective expression and/or over-transcription of similar numbers of other sets of genes, is indispensable for normal regulation, synchronization and progression of consecutive processes of differentiation, specialization, growth and development. Comparative analysis of genome, transcriptome and proteome functional maps during each of these stages of embryonic and fetal life could reveal the intimate correlations between synchronized temporal regulation of these three components and the genomic regulatory aspects of developmental stages and reveal many obscure aspects of this critical period of life. This information might prove helpful in understanding genomic and genetic disturbances that predispose to the pathogenetic mechanisms and the pathophysiological alterations that mediate and underlie the development of congenital malformations. Comprehensive understanding of these aspects of development represents a prerequisite step indispensable for hypothesizing and designing effective prophylactic measures and early therapeutic approaches against teratogenesis and development of congenital malformations. Additionally, formal analysis of databases of structural and functional genomic and proteomic maps of fetal development can add, significantly, to our current vague and scanty knowledge of many, still, unknown aspects of regulatory genetic mechanisms, particularly those of genetic and genomic imprinting and temporal synchronization of mass suppression and activation of huge numbers of genes that characterize development and differentiation.

Constructed integrated formal genetic maps might be helpful for many purposes. Comparative analysis of integrated maps constructed for transcriptome and oncoprotein profiles of pathologically distinct types of malignancies might offer rapid and accurate diagnostic tools of these tumors when compared to corresponding maps of normal cells. Delineation of *tumor maps* comprising the whole spectrum of proto-oncogenes, oncogenes and tumor suppressor genes in the human genome represents an ultimate objective and an indispensable goal crucial for understanding pathogenetic mechanisms underlying the development of cancer. Integration of these maps with oncoprotein maps and their sequential analysis during different stages of tumorogenesis would, undoubtedly, reveal dynamics of tumor progression and acquisition of the different characteristics of the malignant phenotype of cancer cells. Understanding these aspects is crucial for proper designing of therapeutic approaches to cancer based on targeting oncoproteins responsible for the development of the malignant phenotype. The roles played by proto-oncogenes and tumor suppressor genes in regulating vital cellular processes, like the cell cycle and cell division, allow for proposing many hypothetical assumptions regarding their possible therapeutic uses as growth promoters in genetic disorders characterized by defective/deficient/disturbed cell growth/division like some types of short stature. Similarly, the critical roles played by specific sets of oncoproteins in mediating/maintaining/furthering some essential distinctive features of the malignant phenotype, e.g., neovascularization, allows for proposing their use as therapeutic agents in disease states resulting from ischemic injuries to different tissues, e.g., thrombotic strokes, coronary obliterations and diabetic vascular diseases.

Computational analysis of bioinformatics databases represented by different types of formal genetic maps, whether for comparative or predictive purposes, represents a real challenge. Not only because of lack of many pieces of information necessary to complete these maps, but also due to paucity of our knowledge as regards basic essential genetic mechanisms that regulate the structural integrity and functional specialization of the genome as a whole. For instance, the nature of the regulatory systems that master the highly conserved species-specific genetic constitutions of all creatures, including the human genome, is extremely vague, without any clues to the possibility of disclosing the design or the components of the genetic material responsible for it, or the mechanisms of their action, even in the near future. Similarly, the significance of many puzzling and enigmatic phenomena regarding the known detailed structure and actual functions of the genetic material continues to be a matter of debate in spite of the many theories and hypotheses trying to explain them. Examples of such phenomena include genetic imprinting, transposon activity, apoptosis, metabolic adaptations, malignant transformation and many others. However, the mechanism through which solid biomolecules, notably nucleic acids and proteins, can conceive, interpret and execute pre-programed biological processes that are embodied within their structural configuration remains on the top of these unsolved enigmatic phenomena in biology (Table 2).

6. Physical genetic maps

6.1. Genome sequence maps

Structural genomic maps reveal detailed structural organization of both repetitive and non-repetitive regions of the genome. They reveal the details of the structure, sequence pattern, distribution and arrangement of all genic and non-genic components of the whole genome. These maps can be constructed to delineate structural characteristics of any part of the nuclear genome, *nuclear genome maps*, and the mitochondrial genome, the *mitochondrial genome maps*. Genome sequence maps represent the specific base sequence of the DNA of the whole genome. They include all specific features of the DNA as regards the number and arrangement of the bases along each strand, presence and distribution of single

Table 2 Applications of formal genetic maps.

Type of genetic map	Relevance to medical genetics
1. Family pedigree	Genealogical study reveals pattern of inheritance, type of genetic disease, participation of non-genetic factors in pathogenesis of the disease, etc.
2. Structural genomic (DNA) maps	Provide information on number and distribution of unique genic and unique tandem repeats of non-genic parts or domains of DNA, intergenic regions and repetitive sequences, gene density, etc.
3. Functional genomic maps	Provide information on distribution of functional genic sequences, non-functional sequences like pseudogenes and duplicated non-functional genes, transposons, pyknons, mutable hot spots of DNA, imprinting centers, etc.
4. Genomic structural variants (SVs) maps	Reveals human genetic variations and their possible linkage with specific human diseases. Also, they are of particular significance in many fields of study of comparative and evolutionary genetics.
5. Ribonucleic acids (RNA) maps	Constituent databases of structure and function of different types of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA) and different species of small or microRNAs.
6. Transcriptome maps	Provide databases of (mRNA): structural variations, rates of transcription, of translation and of turnover and decay. Differential characteristic alterations of post-transcription modifications in specific disease states. Comparative analysis of mRNA maps can reveal possible causative gene mutations as well as possible underlying pathogenetic mechanisms.
7. Chromosome maps	Reveal details of chromosome topology, e.g. type of chromosome, gene loci, distribution of telomeres and of ribosomal gene repeats, chromatin type and variations, etc.
8. Gene maps	Clarify the size, base sequence, number of exons and introns, distribution of hot spots and of CG sites, sequence type and organization of promoter region of the gene, etc.
9. Proteome maps	Depicts constituent structural and catalytic proteins. Specific proteome maps can be constructed for specific states, e.g. oncoprotein maps of malignant cells. These can further be classified and delineated according to specific tumor type, e.g. oncoprotein map of hypernephroma, multiple myeloma, etc.
10. Linkage maps	Reveal recombination frequency of different types of genetic markers, e.g., genes/traits/proteins/DNA markers, for identifying location of genes relative to each other on chromosomes, etc.
11. Haplotype maps	Provide comparative data about inter-individual single nucleotide polymorphism to determine the likely locations and haplotypes involved in pathogenesis of, or predisposition to, specific diseases, reveal presence and incidence of linkage disequilibrium of certain haplotypes, etc.
12. Inter-species hybridization maps	Allow for assigning or (mapping) specific genes to specific chromosomes and even to certain chromosome segments.
13. Restriction fragment length polymorphism (RFLP) maps	Used for the diagnosis of point mutations that alter a restriction site, and for comparative purposes, e.g. paternity testing.
14. Radiation hybrid maps	Provide data about relative positions of specific genic and DNA markers on specific chromosomal regions based on frequency of chromosomal breakage induced by radiation.
15. Probe-specific maps	Reveal widespread inter-individual as well as inter-species and intra-species genomic and transcriptomic differences, provide critical data for population, comparative, experimental and evolutionary genetic studies and researches.
16. Genotype maps	Correlate specific genotypes with specific disease states, thus providing crucial information relevant to provisional clinical diagnosis, effective prophylactic management and genetic counseling.
17. Phenotype maps	Depicts pathognomonic diagnostic combinations of disease-specific signs and symptoms for genetically-determined and genetically-mediated disorders.

nucleotide polymorphisms (SNP), presence and pattern of distribution of different types of repetitive sequences and additional characteristic structural organization feature of the genome. In disease states, comparison of the patient's DNA sequence map with normal maps can reveal existing point mutations e.g. transitions, transversions, frame shifting, non-sense (truncation) mutations, etc. These maps are used for recognition of genetic diseases caused by point mutations, and as DNA markers in ID testing in forensic medicine.

Nuclear genomic maps reveal specific DNA markers on the gene level as well as on the whole genome level. DNA markers are defined repetitive sequences of the nuclear genome. They might be short or long sequences ranging in size from few nucleotides to thousands of nucleotides, most often located in noncoding regions of the genome. DNA markers have major diagnostic implications since they represent characteristic structural features of specific regions of the genome. They have, also, critical implications in comparative analysis of genomic profiles of different species. DNA markers have proven to be extremely useful for localizing human disease genes that comprise some of these markers within their sequences, or genes located adjacent to, or near, them. Valuable DNA markers include:

Restriction fragment length polymorphisms (RFLP)
Variable number of tandem repeat polymorphisms (VNTRs)
Single nucleotide polymorphisms (SNP)
Long Interspersed Elements (LINE)
Short Interspersed Elements (SINEs).

6.1.1. Restriction fragment length polymorphisms (RFLP)

Restriction fragment length polymorphism sequences (RFLPs) or restriction site polymorphism are defined by the presence or

absence along the DNA of specific sites, called restriction sites, for specific bacterial endonuclease restriction enzymes. These enzymes break apart strands of DNA wherever they contain certain nucleotide sequences which are characteristic of the specific enzyme [**see** Section 8.2 **of this article**].

6.1.2. Variable number of tandem repeat polymorphisms (VNTRs)

Variable number of tandem repeats is due to DNA markers composed of a variable number of nucleotide sequences that are repeated several times in non-coding regions of DNA. In each case, the number of times a sequence is repeated may vary, and each variant acts as an inherited allele, allowing their use for personal or parental identification. VNTRs comprise two main families: *microsatellites* or repeats of sequences less than about 5 base pairs in length and *minisatellites* composed of longer sequence blocks. Applications of maps of the variable number of tandem repeats include their use in population and comparative genetic studies and their analysis for DNA fingerprinting in forensic medicine.

6.1.3. Single nucleotide polymorphisms (SNPs)

Single nucleotide polymorphisms are individual point mutations, or substitutions of a single nucleotide, that do not change the overall length of the DNA sequence in affected regions. SNPs occur throughout an individual's genome in very large numbers. As of July 2013, dbSNP listed 62,676,337 SNPs in humans [1]. Much has been assumed about the significance of SNPs and their prospective uses in personalized medicine and other similar applications. Actually, apart from their use as individual genome-specific markers useful for comparative diagnostic purposes, e.g., DNA fingerprinting, SNPs maps have no other true or useful applications, and statements regarding their prospective roles in genetic medicine are non-sense assumptions ignoring the basic facts about mechanisms underlying pathogenesis of genetic diseases and the roles played by the proteome in their causation. There is much confusion between the presence of SNPs as normal individual, genome-specific, molecular markers and their relation to development of specific genetic diseases, e.g., Alzheimer disease, sickle cell anemia and osteoporosis, because once a SNP results in causation of a disease state it cannot be considered as a normal molecular marker, rather it should be looked at as a pathogenetic point mutation changing a normal codon to an abnormal codon resulting in translation of a defective disease-causing protein, or even total cessation of protein synthesis or synthesis of truncated protein products in the case of changing a functional codon by a SNP into a non-sense codon.

6.1.4. Long interspersed elements (LINE)

These are genetic elements found in the human genome and other eukaryotic genomes. The human genome contains over one million LINEs sequences representing nearly 17–19% of the total genome size. The most abundant of these belong to a family called LINE-1 (L1). These L1 elements are DNA sequences that range in length from a few hundred to a few thousand base pairs. A very minute portion (about 50 L1 elements) are functional genes having sizes about 6500 bp in length and encoding three distinct proteins including an endonuclease and a reverse transcriptase. The sequence diversity of LINEs elements between individual human genomes make them useful markers for DNA finger printing [2].

6.1.5. Short interspersed elements (SINEs)

Short interspersed elements (SINEs) are short DNA sequences, about 100–400 base pairs, found in very high copy numbers in the human genome and many other mammalian genomes. Nearly, 10% (about 300 Mb) of the human genome is composed of a single family of SINEs, known as the *Alu*s. Alu elements consist of a sequence averaging 260 base pairs that contains a site that is recognized by the restriction enzyme AluI. Most SINEs do not encode any functional molecules and depend on the machinery of active L1 elements (reverse transcriptase system) to be transposed, accumulated by a copy and paste mechanism and gets inserted and integrated into the genome at distinct locations. Like the sequences diversity of LINEs elements between individual human genomes, SINES elements are also useful markers for DNA finger printing [2].

6.2. Gene maps

The gene map is a descriptive representation of the detailed structure, or the internal anatomy, of the gene. It reveals the size of the gene or the number of nucleotides it has, the specific base sequence of the whole gene comprising the promoter region/the exons/introns and the terminal region. It delineates also the number and the arrangement of exons and introns along the gene (Fig. 1).

Gene maps are constructed for each gene to reveal the structural organization of the gene which is mandatory for the molecular diagnosis of pathogenetic alterations of gene structure induced by mutations, whether caused by point

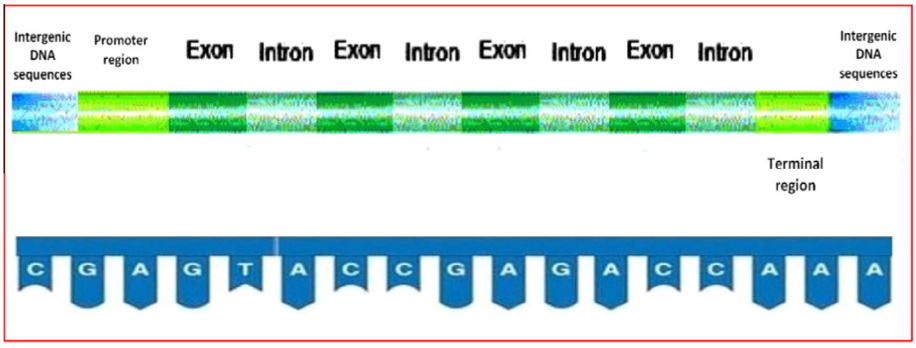


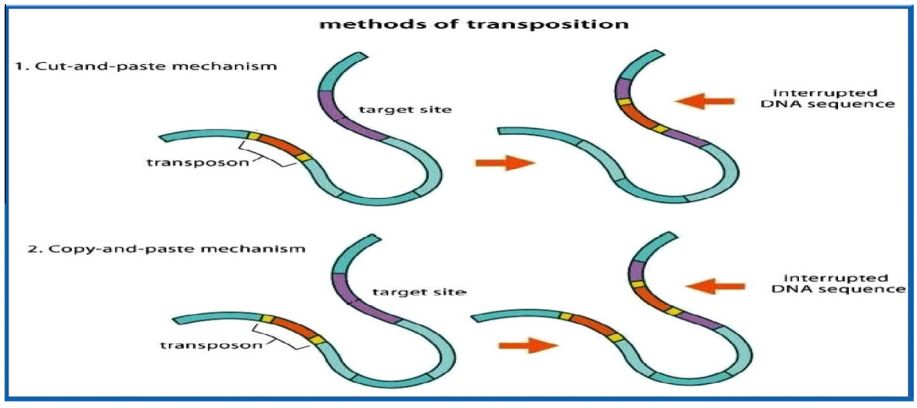
Fig. 1 Schematic representation of the gene map.

mutations affecting one single base of the gene, or small mutations affecting few bases/one or more exons/one or more introns, or even the whole gene in cases of deletions, duplications or rearrangements. Gene maps can be further classified into *exome maps* detailing the number, structural organization, sequence pattern, SNP distribution and functional codon profile of all exons of the gene and *introme maps* detailing the number, structural organization, sequence pattern, SNP distribution and other intronic retrotransposed and repetitive elements of all introns of the gene.

These maps are used for the diagnosis of point, as well as, small-sized and gross mutations of the gene, e.g., single or multiple exon or intron deletions/insertions/inversions/amplifications/improper crossing over during recombination. Point and few nucleotide mutations are amenable to diagnosis by gene sequencing techniques. Large mutations, on the other hand, can be diagnosed by many techniques that depend on the integrity of the base complementarity of the gene, e.g., Fluorescent In Situ Hybridization (FISH). Exome maps can be constructed and used for both intra and inter-species comparative purposes. Similarly, comparison of exome maps of patients with specific genetic disorders with those of normal subjects might prove to be a promising approach that can have many diagnostic applications in clinical genetics. Comparative analysis of specific exome maps of specific disease genes can pinpoint the pathogenetic alteration(s) underlying the causation of the disease thus revealing the cause of the disease, the possible mechanisms involved in its pathogenesis and might be useful in proposing and designing effective/alleviating therapeutic approaches to the disease.

6.3. Transposons maps

Transposons or transposable elements (TEs) (Fig. 1), comprising nearly half of the human genome, are genomic sequences found within the introns of most genes. While nearly all TEs within introns appear harmless, some *de novo* intronic TEs insertions do disrupt gene transcription and splicing and cause disease. It is unclear why some intronic TEs perturb gene transcription whereas most do not. Transposons have the ability to detach themselves from their genomic sites to be inserted at other sites in the genome. Alternatively, transposon sequences might make a copy of their selves and this copy gets inserted at another site along the genomic DNA [3]. These findings suggest that transposon activity may represent an important cause of spontaneous genomic mutations capable of inducing genetic damage and instability of the genome. The role played by transposons or transposable elements in inducing certain types of spontaneous mutations, transposition or insertional mutations, is well known in bacteria [4].



Transposons: methods of transposition (Lauren Solomon, Broad Institute of Harvard and MIT).

The role played by increased *transposon activity* in causing, seemingly, *spontaneous mutagenic events* that predispose to developmental malformations during fetal development deserves more attention if effective prophylactic and therapeutic anti-teratogenic measures are to be reasonably hypothesized and properly designed. Also, the effective opposition of this teratogenic mechanism by the silencing action exerted by P-element induced wimpy testis, Piwi-interacting RNAs (piwiRNAs) or piRNA, which is the largest class of small non-coding RNA molecules expressed in animal cells [5] on transposon regulatory sequences, represents an important regulatory mechanism responsible for maintaining genomic integrity and genomic stability during this critical period of life. Construction of structural maps of transposons (transposon maps) seems mandatory because comparison of *malformation-specific transposon maps* with their normal counterparts would offer invaluable information which can be used to construct anomaly-specific databases correlated with these particular transposon maps that depict specific overexpressed transposons in specific types of congenital malformations. These databases could have predictive value in delineating *malformation-causing transposons* and might prove useful for prenatal diagnosis of these malformations. Further, they might also prove to be important as potential therapeutic/prophylactic applications by choosing and designing proper and specific piRNA silencing molecules against these particular anomaly-causing transposons.

6.4. Pyknon maps

Pyknons (Fig. 2) are short DNA sequences, about 20-22 nucleotides in length, widely distributed in both the inter-genic and intronic regions of the nuclear genome of human as well as in the genomes of many other organisms, e.g., fruit fly, chicken, mouse, rat, and dog, the numbers of found human pyknons in other species decrease with their phylogenetic distance. Additionally, more than 90% of all human genes contain one or more pyknon instances. More than 200.000 copies/patterns of non-overlapping pyknons comprising nearly 900 million nucleotides of the human genome constitute about *1/6th* of the human genome. This makes them the most frequent, variable-length DNA sequence motifs in the human genomes. Pyknons have a remarkable degree of structural conservation. Their presence in the 3′ UTRs (un-translated regions) of genes may indicate a potential regulatory role in posttranscriptional processing and modifications of mRNA. Though they do not share in either protein synthesis or RNA transcription, pyknons are functional genetic elements associated with the mediation of specific biologic cellular processes. They are putative factors implicated in susceptibility to some common human genetic disorders. Disturbed genomic regulation of function(s) of pyknons might underlie the development of this genetic susceptibility [6]. The exceedingly large numbers of variable-length pyknon patterns, their remarkable degree of structural conservation, their widespread distribution in the genome, their existence within coding genes, their putative roles in regulation of many cellular functions and in conferring susceptibility to many genetic disorders would all make construction of pyknon maps a worthy task for many reasons. *First*, they can be used as individual-specific DNA markers for comparative purposes, e.g., paternity testing and confirmation of genetic identity. *Second*, detection of, and correlation between, characteristic patterns of pyknons in specific diseases might offer a predictive/diagnostic approach to these diseases. *Third*, the presence of pyknons in the genomes of other species in varying proportions related to the phylogenetic diversity of these genomes make them a valuable tool in comparative and evolutionary genetic studies.

6.5. Pseudogene maps

Pseudogenes are dysfunctional altered copies of normal genes. With very few exceptions, pseudogenes lack protein-coding ability and their products are no longer expressed in the cell. Though pseudogenes have varying degrees of sequence homology to known genes, the true mechanisms underlying their development are not clear, however they might result from the accumulation of multiple mutations within a gene whose product is not required for the survival of the organism. Transfer of fragments of mitochondrial DNA to the nucleus, which is a continuous and dynamic process particularly during instances of DNA repair, and their incorporation as non-encoded sequences in the nuclear genome account for the origin of a particular class of oncogenes referred to as nuclear mitochondrial pseudogenes (NUMT-pseudogenes) [7] (Fig. 3).

There are four main types of pseudogenes, all with distinct mechanisms of origin and characteristic features: processed or retrotransposed pseudogenes, non-processed or duplicated pseudogenes, disabled genes or unitary pseudogenes and nuclear mitochondrial pseudogenes (NUMT-pseudogenes). Although processed pseudogenes do not have introns or promoters, where they are copied from mRNA and incorporated into the DNA, most other pseudogenes have some gene-like features such as presence of promoters, CpG islands and splice sites. Pseudogenes differ from normal genes in lack of protein-coding ability due to a variety of mutations affecting their promoter regions and lack of transcription of mRNA. The DNA of pseudogenes may be functionally similar to other kinds of non-coding DNA which can have a regulatory role as evidenced by the finding of a retrotransposed pseudogene whose transcript purportedly plays a trans-regulatory role in the expression of its homologous gene. This finding suggests the possibility of mediating important biological regulatory functions by some pseudogenes [8]. This possibility is further supported by the identification of some endogenous siRNAs that appear to be derived from pseudogenes suggesting a role played by these pseudogenes in regulating protein-coding transcripts [9]. Additionally, mRNA levels of the tumor suppressor PTEN gene and the oncogenic KRAS are affected by their pseudogenes PTENP1 and KRASP1, respectively. This discovery demonstrated the important functions of transcripts of these pseudogenes as biologically active factors in tumor biology; thus attributing a novel biological role to expressed pseudogenes, as they can regulate coding gene expression, and reveals a non-coding function for mRNAs in disease progression [10]. Of utmost importance, however, is the finding that a processed pseudogene called phosphoglycerate mutase 3 (PGAM3P) actually produces a functional protein [11]. These

Fig. 2 Pyknons (www.junkdna.com).

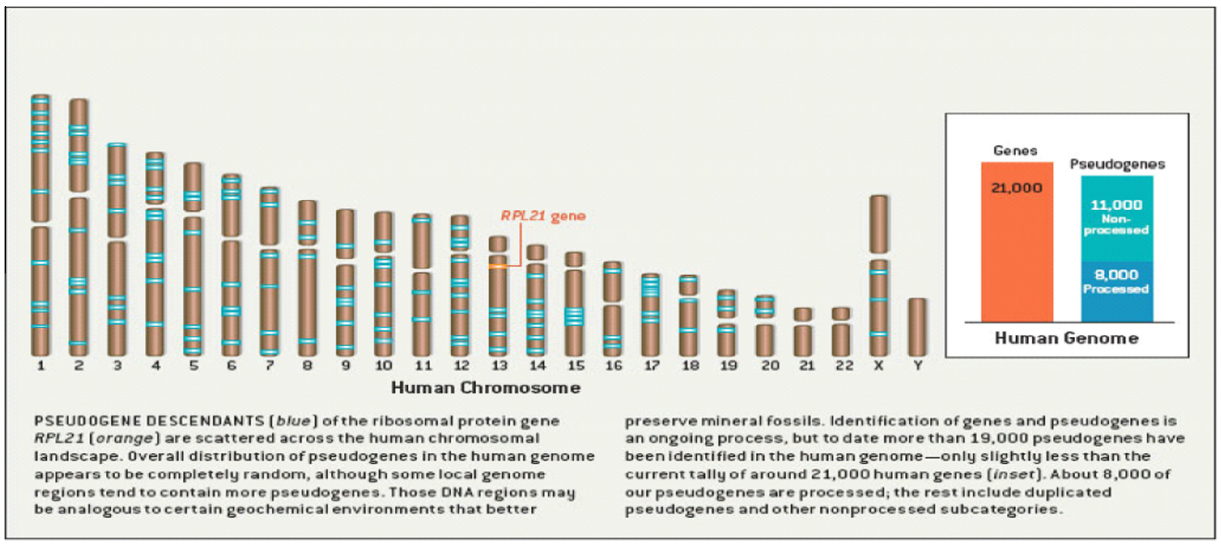


Fig. 3 Pseudogenes in human genome (Mark Gerstein and Deyou Zheng: The real life of pseudogenes, 2006).

findings opposing the presumed non-functionality of pseudogenes raise many queries regarding the true significance of the presence of exceedingly large numbers of pseudogenes, that may nearly equal the number of functioning genes, in the genome and their global distribution across all chromosomes.

Pseudogenes can complicate molecular genetic studies. For example, amplification of a gene by polymerase chain reaction (PCR) may simultaneously amplify a pseudogene that shares similar sequences. Also, nuclear copies of mitochondrial DNA (mtDNA) can contaminate PCR-based mitochondrial studies. This is known as PCR bias or amplification bias. Similarly, pseudogenes are sometimes annotated as genes in genome sequences. In addition, processed pseudogenes might pose problems for gene prediction programs often being misidentified as real genes or exons. It has been proposed that identification of processed pseudogenes can help improve the accuracy of gene prediction methods [12].

Construction of detailed maps of pseudogenes including their types, distribution across the genome and their detailed sequencing is important for many reasons:

First, in view of the problems ordained by pseudogenes in diagnostic/predictive research studies these maps might be helpful, as reference guides, to avoid or minimize confusing/misleading results in this respect.

Second, the possible regulatory roles attributed to some pseudogenes, that might be shared by more pseudogenes or even be a common characteristic of most pseudogenes, necessitate serious consideration of their roles in regulating gene expression of normal genes including oncogenes and tumor suppressor genes.

Third, the global genomic distribution of the exceedingly large number of pseudogenes over all chromosomes contradicts theoretical assumptions as regards the origin/non-functionality of pseudogenes, as well as their unjustified classification as junk portions of the genome, because there is no point in keeping such a considerable portion of non-functional components in the genome. Construction of functional maps of pseudogenes in the human genome would, certainly, reveal still undefined biological functions possibly mediated/affected/regulated by transcripts or even protein products of some pseudogenes.

Fourth, the development and formation of nuclear mitochondrial pseudogenes (NUMT-pseudogenes) entails insertional mutagenic changes of the genome. Affection of functioning genes by these pathogenetic changes can induce structural disruption with consequent loss/suppression of function(s) of these genes. The ensuing pathobiochemical/pathophysiological alterations secondary to deficiency of gene product(s), whether being structural proteins/catalytic proteins/regulatory RNA or other products, can mediate the pathogenesis of genetic disorders.

Development of diseases assumed to be related to, or attributed to, regaining function or dysfunction of some pseudogenes might occur and progress according to this hypothesized pathogenetic mechanism. Accordingly, construction of integrated structural/functional maps of pseudogenes might prove helpful in revealing and detecting some of the mechanisms underlying the pathogenesis of these diseases.

6.6. Telomere maps

Telomeres are repeating sequences of DNA found at the ends of chromosomes of most eukaryotic organisms. They are composed of arrays of guanine-rich, six- to eight-base-pair-long repeats, TTAGGG, varying greatly in number between species, from approximately 300 base pairs in yeast to many kilobases in humans. Eukaryotic telomeres normally terminate with a 3′ single-stranded-DNA overhang, which is essential for telomere maintenance and capping. Telomeres have two main functions in the genome. *First*, they prevent the ends of the chromosome from fusion with other chromosomes thus maintaining the chromosome number of the cell, which is a prerequisite for proper cell division. *Second*, they protect the ends of the chromosomes from progressive shortening which occurs with cell division because DNA replication does not continue to the end of a chromosome. Hence, telomeres are replenished by a telomerase enzyme system and compensate for incomplete semi-conservative DNA replication at the chromosomal ends. Mammalian somatic cells lacking telomerase gradually lose telomeric sequences as a result of incomplete replication. As the telomeres get shortened, the cell eventually reaches its rep-

licative limit and progress into senescence or old age through degenerative pathways involving the p53 and pRb pathways leading ultimately to halting of cell proliferation and eventual cell death [13]. Considered together, these functions of the telomeres are, undoubtedly, indispensable for preserving species-specific genomic identity, integrity and stability.

In view of these findings, defective/deficient functions of telomeres are expected to induce a wide spectrum of pathogenetic alterations of affected cells including enhanced senescence and aging, increased incidence of chromosomal rearrangements and the development of cancer. A common finding among different types of tumors is the presence/maintenance of telomerase activity in tumor cells to compensate for the ongoing telomere shortening due to the rapid pace of cell division of malignant cells. These cells use telomerase to extend telomere lengths, allowing the tumor to continue its rapid growth [14]. In light of this finding, abolishment of telomerase activity via silencing of the telomerase gene might prove to be an effective therapeutic approach to combat oncogenesis. However, it must be realized that proper regulation of telomerase activity should be maintained since suppression of telomerase activity would result in enhanced shortening of telomeres leading to premature aging and senescence. On the other hand, persistent expression of telomerase for too long bears the risk of increased incidence of tumor formation.

Construction of integrated telomere maps for specific cell types comprising telomere lengths determined by e.g., quantitative telomeric FISH technique, state of telomerase activity and state of chromosomal stability identified by high resolution banding might be helpful as predictive parameters for anticipating the development of cancer in predisposed individuals having persistently maintained telomere lengths, high expression levels of telomerase activity and chromosomal rearrangements commonly associated with tumor development, e.g., increased incidence of chromosome breaks. Integration of telomere maps with corresponding oncogene maps and, more significantly, with maps of oncoproteins would, probably, add more relevance to these predictive approaches to carcinogenesis.

However, assumptions that attribute immortalization of cancer cells and their extended proliferative potential and survival primarily to their ability to maintain high expression levels of telomerase with corresponding maintenance of telomere lengths of cancer cells [15] and postulations regarding roles of telomeres in the regulation of gene expression and differentiation of cancer cells [16] in addition to hypotheses relating telomere lengths to aging, senescence, and development of age-dependent pathological states [17] all ignore the critical roles of factors, other than telomeres, directly implicated in initiation and progression of age-related degenerative genomic and cellular changes. These factors include, for example, the predefined programed apoptotic systems, the progressive mutagenic burden of both the nuclear and the mitochondrial genomes caused by continuous exposure to the damaging effects of endogenous and exogenous mutagens, and the corresponding intemperate deteriorating effects of these changes on structure/organization/function of nuclear and cellular proteome compartments responsible for maintaining pivotal proper structural integrity and optimal functional potentials of nuclear/mitochondrial/cellular components. Damaging and degenerative effects of these factors include direct mutagenic damage of genes responsible for regulating DNA replication/cell growth/cell division, direct mutagenic damage of tumor suppressor genes, direct mutagenic damage of genes responsible for synthesis of proteins that mediate and regulate oxidative-phosphorylation networks responsible for ATP production, and degenerative/lethal alterations of structural/intermediary/metabolic networks responsible for regulating and maintaining vital cellular processes like signal transduction/trans-membrane transport/stability of cytoskeleton/cell contact and interaction dynamics/protein modification and trafficking/cellular secretory functions and many others. It is obvious that any of these changes that are not dependent on, or related to, telomere structure or function can have lethal and drastic effects on the cell even to larger extent than defective telomere structure/function can induce. They can induce arrest of DNA replication and impairment of cell division, direct cell death, development of cancer, aging and senescence due to progressive failure of cellular functions, defective structural organization of cell components and defective/deficient regulation of vital cellular functions. Unjustified over exaggeration of postulations regarding the roles of telomeres as guardians of cell life might mask and/or underestimate the equally, or even more, important roles of other genomic components in maintaining life processes in the cell.

6.7. Chromosome maps

Chromosome maps reveal detailed structural organization of different regions of the chromosomes. They comprise primarily gene locus maps that define spatial localization of genes on chromosome arms/bands/sub-bands/sub-sub-bands (Fig. 4). Maps of chromosomal constitutional polymorphic markers are relevant to studies of comparative and evolutionary genetics since detecting striking variations of the most common of these variants or cytogenetic markers, e.g., peri-centric inversion of chromosome 9, polymorphisms of size and peri-centric inversion of Y chromosome and polymorphisms (very small/deficient) of the heterochromatic regions of chromosomes, revealed that the incidence of specific chromosomal variants is different in each population group [18]. The significance of these maps of chromosomal polymorphic variants stems also from findings pointing to role(s) of chromosomal inversions in the development of adaptation, promotion of speciation and reproductive isolation in natural populations [19].

Mapping of specific genes to particular chromosome regions has been achieved through many approaches. Observation of sex differences in trait/disease phenotype is the simplest of these approaches as it broadly classifies genes into autosomal and sex linked, X-linked and Y-linked, genes. Detection of a persistent linkage between absence or presence of specific chromosomal regions with the absence or presence of specific gene products, e.g., proteins/immunoglobulins/hormones etc., e.g., chromosomal deletion studies, has been used to exclude or verify mapping of genes encoding these products to these chromosomal regions, respectively. Somatic cell hybridization experiments or inter-species hybridization mapping (**see** Section 7.1) had the greatest impact on constructing detailed chromosomal maps of nearly each chromosomal segment of the human genome and resulted in formulating comprehensive maps of most gene loci located on chromosomes. Additional approaches used for mapping genes and constructing chromosomal maps included linkage studies that indicate approximate distances between genes on specific chromosomal regions,

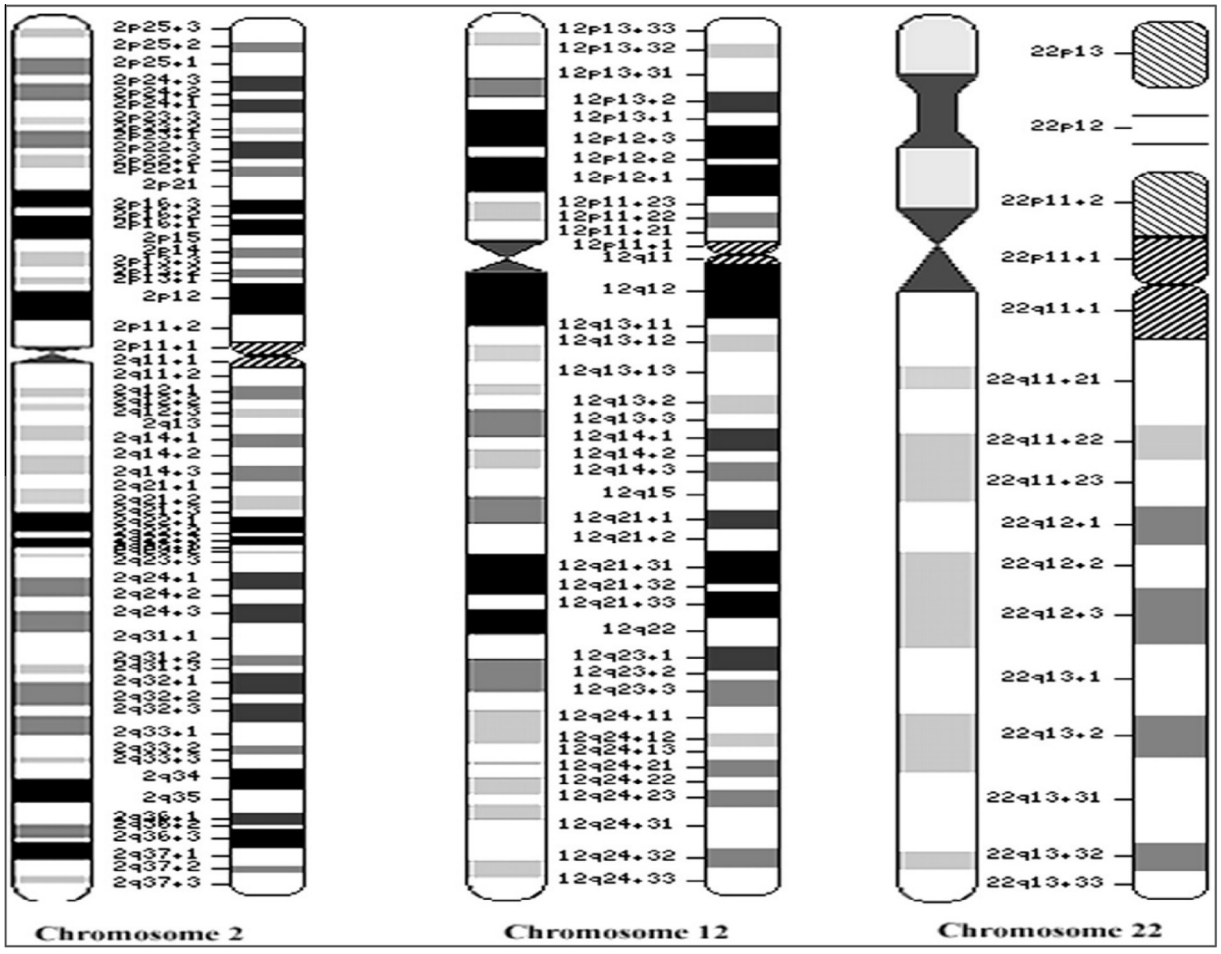


Fig. 4 Gene locus maps of some chromosomes (www.soe.ucsc.edu/research/compbio/cytobands/).

marker-linked mapping techniques that predict mapping of specific idiopathic genes to specific regions already having genes including these markers, e.g., structural DNA markers, or known to be responsible for coding these markers, e.g., biochemical markers. Researches aiming at the detection of hybridization and complementation between whole unprocessed primary, or heterogeneous nuclear, mRNA molecules to denatured single stranded genomic DNA might be helpful in detection of putative genes responsible for transcribing these mRNA molecules and in mapping them to specific sequences of the genome.

7. Functional genetic maps

The significant and beneficial applications of structural genetic maps in different fields of genetics including medical genetics call for construction of parallel functional genetic maps that characterize critical functional regions and transcriptionally active domains of the human genome. Hypothetical applications of functional genomic maps can encompass a wide range of prophylactic, diagnostic and therapeutic applications. For instance, anticipation of possible complications in patients suffering from specific genetic disorders, and of possible health hazards in individuals with susceptible genetic backgrounds based on formal genetic maps or databases of their genomic and proteomic constitution, represents an important prophylactic approach with favorable prognostic prospective in the management of patients with specific genetic diseases and in offering proper counseling advice to their concerned family members.

Functional genetic maps can be constructed for any functional region of the genome. As stated previously, the differentiation between physical or structural genetic maps and functional genetic maps is not absolute Considerable overlap between both types of maps do exist and some types of genetic maps can be considered within the context of both groups. Examples of vital functional genetic maps include exome maps, transposon maps, pseudogene maps, pyknon maps, telomere maps, transcriptome maps and proteome maps.

7.1. Transcriptome maps

On the molecular level, comparative analysis of transcriptome maps of messenger RNA and of different micro RNA species in normal and in diseased subjects could offer remarkable help in the diagnosis of genetic diseases characterized by substantial genomic activities. Malignant cells express large numbers of actively transcribed genes and detailed *cancer-specific transcriptome maps* could be constructed for pathologically distinct types of malignancies, thus allowing for better understanding of the aberrant phenomena of differential suppression of certain genes and of mass overexpression of specific gene families, e.g., proto-oncogenes, in malignant cells. They can also offer

relevant information on the nature of different pathogenetic mechanisms underlying the development of specific types of malignancies as well as on the dynamics of tumor progression based on comparative quantitative assays of tumor-specific transcriptome maps and oncoprotein maps.

7.2. Proteome maps

Proteome maps comprise and delineate detailed information databases of the whole set of proteins encoded by the whole genome as regards: (1) *Nature* of protein components whether they are *structural proteins* like cell cytoskeleton proteins, *catalytic proteins* like enzymes or *intermediary proteins* like signal transducing proteins. (2) *Source* of protein components whether encoded by nuclear genes or mitochondrial genes (3) *Spatial localization* in different compartments of the cell e.g., in cell membranes, cell pores, cell organelles, mitochondria, cell nucleus, DNA-associated chromatin etc. (4) *Function* in different cellular processes e.g., cell cycle regulating proteins, membrane transport regulating proteins, signal transduction mediating proteins, lysosomal/peroxisomal enzymes etc.

Attempts aiming at construction of comprehensive proteome maps comprising the whole set of proteins in the cell would, probably, have far reaching beneficial impact on many aspects of therapeutic genetics. For instance, comparative analysis of proteome maps in normal and in subjects affected with specific genetic disorders can identify disease-specific defective/deficient proteins responsible for pathogenesis of these disorders. This diagnostic approach might prove particularly helpful for idiopathic genetic diseases, resulting from deficient synthesis of proteins or synthesis of defective proteins, the causative genes of which are not yet identified. It can also help in designing specific treatment approaches based on detected proteomic changes, e.g. offering deficient/defective proteins to patients affected with these disorders.

Use of proteome maps in identifying still-undetected genes, similar to reverse engineering techniques, might be worthy of trial as a possible gene mapping approach. Computational construction of possible mRNA codon sequences that can be translated to specific proteins, combined with computational construction of possible corresponding nucleotide sequences that can be transcribed to the specific mRNA in question might help in defining most of the exome sequences of transcribing genes in DNA sequence maps. Differentiation between coding and non-coding sequences of DNA is a prerequisite before matching between both sequences due to presence of intronic sequences within the gene. Synthesis of a complementary strand of the postulated exome sequence and its use as a hybridization probe might indicate, in a more or less approximate manner, possible localization/mapping of the putative gene. Additionally, computational analysis of correspondence between both sequences may be helpful in detecting the most possible locus among detected loci.

7.3. Oncoprotein maps

Comparative analysis of proteome maps of normal cells and of malignant cells can delineate and depict functional deviations of the genome in cancer cells and detect oncogenic proteins synthesized and expressed by the malignant cell. Construction of oncoprotein maps comprising proteins responsible for pathogenesis and evolution of the malignant phenotype would, certainly, have a beneficial impact on research studies aiming at better understanding of malignant transformations through revealing pivotal oncoproteins that initiate and promote the malignant phenotype of cancer cells. They would, also, have a similar impact on research studies aiming at designing and tailoring specific therapeutic approaches for challenging carcinogenesis through e.g., engineering of monoclonal antibodies against relevant oncoproteins inside intracellular compartments as well as on cell membranes and within intercellular spaces/adhesion domains of malignant cells. Interruption of signaling pathways involved in initiation/mediation/regulation of selective advantageous characteristics and over expressed functions of malignant cells, e.g., enhanced cell division/suppressed apoptosis/enhanced glycolysis/neovascularization/metastasis, through disassembly of metabolic/regulatory/intermediary networks actively participating in malignant transformation and evolution of the malignant phenotype could, probably, represent a promising approach in proposing and designing effective treatment modalities of cancer.

7.4. Regulatory networks maps

As all life activities in cells are mediated by a countless number of regulatory networks, each comprising tens to hundreds of proteins and other non-protein molecules, delineation of these networks and their constituent protein components and construction of formal maps detailing their structure/function relationship would be a major advance in understanding pathogenetic mechanisms underlying disease pathogenesis and revealing the dynamics of pathophysiological alterations of different stages of development of the disease phenotype. Construction of regulatory networks involved in the modulation of gene functions, apart from networks regulating proteome functions, would also have a great impact on understanding mechanisms of gene activation and factors involved in gene repression.

Recent studies aiming at generation and construction of a genome-wide interaction map of regulatory elements in human cells revealed that interactions between regulatory genomic elements play an important role in regulating gene expression. The study also revealed that comparison of interactions between cell types shows that enhancer–promoter interactions are highly cell-type specific, thus revealing new mechanistic and functional insights into regulatory region organization in the nucleus [20]. The results of this study reveal the importance of constructing maps of gene functions detailing the different factors responsible for regulating gene function, whether through activation or suppression. Assumptions regarding possible uses of these factors to regulate and control gene functions in treatment/alleviation/prevention of genetic disorders, seem quite reasonable compared to classic gene therapy techniques that drastically affect targeted genes in view of their indiscriminate haphazard action and the major risks of resulting insertional mutagenesis.

8. Induced/experimental maps

8.1. Inter-species hybridization maps

The observation that genomes of some organisms have preferential selective advantages over genomes of other organisms

when allowed to mix together *in vitro* had crucial implications on early attempts for mapping specific genes to certain chromosomes. When somatic cells of two different species are mixed together and exposed to the effect of certain cell membrane fusing viral vectors, e.g., Sendai virus 40, in cell culture media they combine together forming a single cell with a common fused nucleus called hybridoma. Somatic cell hybridization experiments between different organisms revealed that inter-species genomic selective advantage results in consequent progressive loss of chromosomes of one organism with persistence of the chromosomes of the other organism. Accordingly, monosomic hybridomas for specific chromosomes with proteome profiles defined by the remaining chromosomes of the hybridoma, result. Biochemical characterization of proteins expressed by a distinctive hybridoma with unique chromosomal constitution combined with comparative analysis and correlation of presence/absence of specific proteins with presence/absence of certain chromosomes of the hybridoma allows for assigning/localizing/mapping of specific genes to specific chromosomes and even to certain chromosome segments persistently expressing proteins encoded by these genes in the culture media [21]. Though inter-species hybridization maps constructed based on comparison/correlation analysis of inter-species hybridoma experiments provide a crude technique for gene localization and mapping of specific genes to specific chromosomes, they have great impact on gene mapping techniques as a prelude to delineate and construct chromosome maps of, still, unlocalized genes (Figs. 5 and 6,).

8.2. Restriction fragment length polymorphism maps (RFLPs)

Synthesis of enzymes capable of breaking DNA is one, among many, defense mechanisms whereby bacteria can damage infecting viruses. These enzymes, or *endonucleases*, attack specific sequences present on double stranded DNA so that whenever these sequences, or *restriction sites*, exist along the DNA molecule they got broken by the enzyme with the ultimate result of fragmenting the target DNA into many fragments that have peculiar numbers, sizes and distributions defined by the type of the endonuclease enzyme, the number of restriction sites, their sequence specificity and their distribution along the DNA strand. Restriction fragment length polymorphism maps are constructed by digestion of DNA by specific endonucleases which target their specific restriction sites in the DNA causing its break into a DNA–endonuclease peculiar breakage pattern (Fig. 8). The resulting characteristic breakage pattern, or fragment length polymorphism map, have many applications, for instance, they can be used for diagnosis of genetic diseases caused by point mutations that alter the sequence of a restriction site leading to generation of a different breakage pattern, e.g., sickle cell anemia. They can also be used for comparative purposes, e.g., for paternity testing [22].

8.3. Radiation hybrid maps (RH)

The mutagenic effects of particulate radiations on DNA leading to its breakage and fragmentation have been used to generate specific radiation-induced fragmentation pattern maps based on the positive correlation between radiation dose and DNA breakage index in human–rodent somatic cell hybrids (Fig. 7). The differential susceptibility of chromosomes of each cell type to the delivered radiation dose results in breakage of the human chromosomes into fragments each having one or more specific DNA markers. These fragments are subsequently integrated into the rodent chromosomes. From these human–rodent or donor–recipient hybrids, clones can be isolated and tested for the presence or absence of DNA markers on the human chromosome of interest, and the frequencies with which markers were retained in each clone can be calculated. The idea is based on a statistical method to determine not only the distances between DNA markers but also their order on the chromosomes. The concept of mapping is similar to the underlying principle of mapping genes by linkage analysis based on calculation of recombination events where the farther apart two DNA markers are on a chromosome, the more likely a given dose of X-rays will break the chromosome between them and thus place the two markers on two different chromosomal fragments. The order of markers on a chromosome can be determined by estimating the frequency of breakage that, in

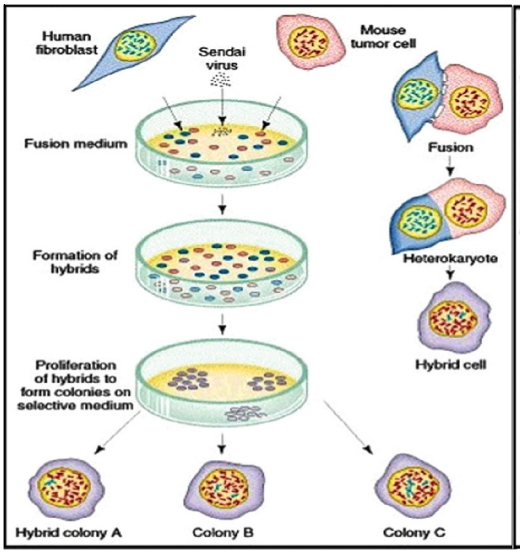


Table 8-2. Mapping a Human Gene by Using Somatic Cell Hybrids

Hybrid	HEXA Gene	Human Chromosomes 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
I	+	+	−	−	−	−	−	−	−	+	+	−	−	+	−	+	−	−	+	−	+	−	−	+	−
II	−	+	+	−	+	−	−	+	+	−	−	−	+	−	+	−	−	−	+	+	−	+	−	+	−
III	+	−	−	−	+	−	−	+	+	−	−	+	−	+	+	+	−	−	+	−	−	+	+	+	+
IV	−	−	−	−	−	−	−	+	+	−	+	−	+	+	+	−	−	−	+	+	−	+	−	+	−
V	−	−	+	+	−	+	−	−	+	−	−	+	−	+	+	−	−	+	−	−	+	+	+	+	+
VI	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	+	−	+	−	+	−
VII	+	+	−	+	−	−	−	+	+	−	+	+	+	−	+	+	+	−	−	+	+	−	−	+	−
VIII	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	+	−

Figs. 5 and 6 Inter-species hybridization mapping technique.

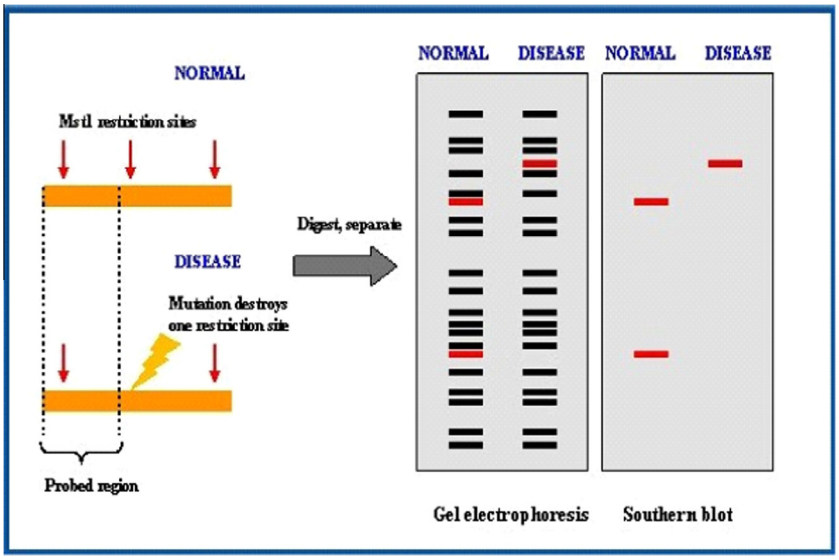


Fig. 7 Technique of RFLP mapping (NCBI–NLM MeSH database resources).

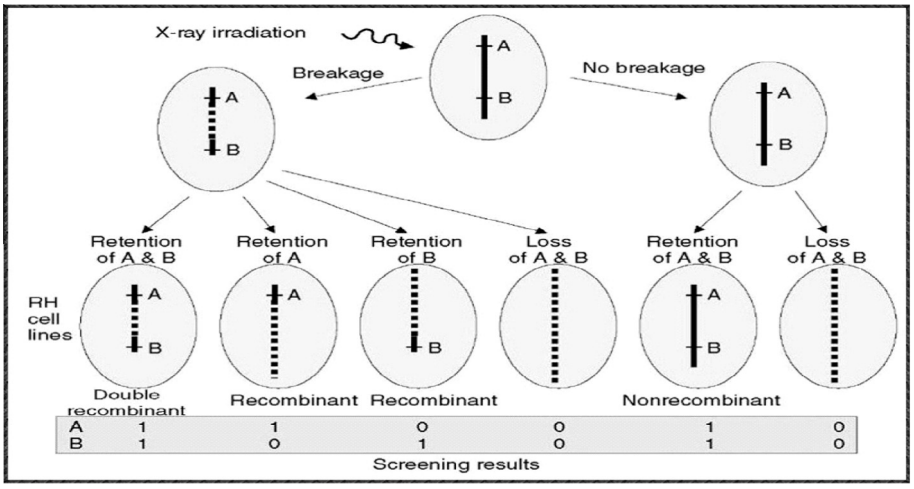


Fig. 8 Principle of radiation hybrid mapping. (www.what-when-how.com/genomics).

turn, depends on the distance between the markers. This technique has been used to construct whole-genome radiation hybrid maps [23].

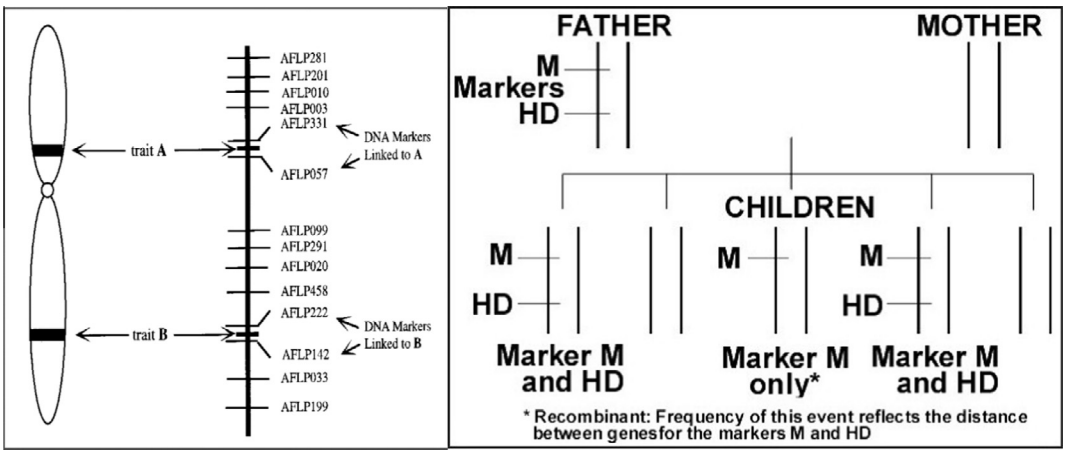
8.4. Probe specific maps

These maps are constructed by hybridizing DNA to a large panel of different probes prepared for multiple complementary genomic markers. These maps can be applied to reveal widespread inter-individual as well as inter-species and intra-species differences, so they are critical for population genetic studies, comparative genetic studies, experimental genetic researches and evolutionary genetic studies. Probe specific maps might prove useful in molecular diagnosis of genetic defects caused by small/large mutations affecting marker sequences of the genome. Comparative analysis of probe-specific maps of DNA of normal and diseased individuals can reveal mutated marker-including regions of the genome. They might also prove applicable for detection of chromosomal rearrangements that result in physical interruption of specific marker sequences involved in the rearrangements, thus leading to generation of a different probe specific map depending on the chromosomal location of the marker(s) and the nature of the chromosomal rearrangement [24,25].

9. Constructed genetic maps

9.1. Linkage maps

A linkage map is a genetic map of a certain species or experimental population that shows the position of its known genes or genetic markers relative to each other in terms of recombination frequency, rather than as specific physical landmarks along each chromosome (Figs. 9 and 10). The greater the frequency of recombination or segregation during crossover of homologous chromosomes between two genetic markers, the



Figs. 9 and 10 Principles of linkage mapping techniques.

farther apart they are assumed to be. Conversely, the lower the statistical frequency of recombination between the markers, the smaller the physical distance between them. The genetic markers that are used for constructing linkage maps for specific wild type normal genes or disease genes include different types of markers, e.g., clinical markers (eye color), genomic markers (minisatellites), biochemical markers (enzymes and proteins), chromosomal markers (telomere maps), and molecular markers (single nucleotide polymorphism).Genetic linkage maps are not true physical maps, rather, they help researchers to locate other markers, such as other genes, by testing for genetic linkage of the already known genetic markers. Linkage mapping is critical for identifying the location of genes that cause genetic diseases [26].

9.2. Genotype/haplotype maps

Genotype maps define the state of alleles that comprise the genetic constitution of the individual. They include maps of bi-allelic as well as poly-allelic traits and diseases. Genotype maps reveal important data relevant to many aspects of basic and clinical genetics. For instance, they comprise allele frequencies, single nucleotide polymorphism (SNPs) and intronic retrotransposed and repetitive sequences of each allele. Allele frequency mapping is crucial for studies of population genetics and genetic linkage studies, and SNPs are used for comparative purposes. Genotype maps of patients suffering from specific diseases might have accurate diagnostic implications when compared with normal maps for the same candidate disease genes. They also have important prophylactic applications in pre-implantation diagnosis and in prenatal diagnosis as well. Comprehensive genotype maps comprising the whole nucleotide sequence and the characteristic structural variants of both, or all, alleles of the gene represent a promising and reliable approach for early detection and diagnosis of subjects with susceptible genetic backgrounds, based on peculiar features of their genetic constitution as detailed in their genotype maps.

9.3. Phenotype maps

Contrary to traditional restricted practices that confine the phenotype of a specific trait/disease to the clinical features observed and detected, informative phenotype maps can be constructed and formulated to include comprehensive representation of relevant data of all possible phenotypic aspects of specific genetic disorders. Thus, broad spectrum phenotype maps constructed by integrating abnormal clinical findings, family history, chromosomal aberrations, defective/deficient biochemical or proteome constitution, transcriptome alterations and genome-wide molecular profile(s) of specific diseases, represent accurate and comprehensive disease-specific maps that can offer reliable diagnostic approaches to these disorders.

9.4. Disease association maps

The persistent significant association of specific disease states with particular genetic markers could be interpreted in many ways. *First*: participation of detected markers in pathogenesis of the disease or in conferring susceptibility to its development. *Second*: in the case of protein markers, the gene(s) causing the disease and the marker-encoding gene may be affected/regulated by the same factors (enhancers/suppressors) that control gene expression. *Third*: tight physical linkage due to proximity of loci on the same chromosomal region. *Fourth*: in the case of sequence markers, e.g., pseudogenes or repetitive sequences, the disease gene(s) might be physically related to the marker sequences or regulated in a positive or negative manner by the non-protein transcripts of these sequences. *Fifth*: association of disease phenotype with particular non-related normal phenotypic features, or abnormal phenotypic features of other disease states, might indicate sharing exposure to developmental regulator networks or synchronized affection by development-disturbing processes like increased transposon activity or exposure to teratogenic agents, respectively.

Accordingly, construction and analysis of comprehensive disease association maps could have a beneficial impact on researches aiming at clarifying the different and relevant aspects of some of these aforementioned theoretical assumptions. Possible use of disease association maps as preliminary predictive/diagnostic tools, through comparative analysis of corresponding maps in normal individuals, and for mapping unidentified disease gene(s) based on prior knowledge of mapping details of associated markers adds more merit to their application in many fields of basic and clinical genetics.

References

[1] NCBI – National Center for Biotechnology Information, United States National Library of Medicine. 2013. NCBI dbSNP build 138 for human.

[2] Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 1999;9(6):657–63.

[3] Zhang Ying, Romanish Mark, Mager Dixie. Distributions of transposable elements reveal hazardous zones in mammalian introns. PLOS Comput Biol 2011. http://dx.doi.org/10.1371/journal.pcbi.1002046.

[4] Kidwell MG. Transposable elements. In: Gregory TR, editor. The Evolution of the genome. San Diego: Elsevier; 2005. p. 165–221.

[5] Seto AG, Kingston RE, Lau NC. The coming of age for Piwi proteins. Mol Cell 2007;26(5):603–9.

[6] Rigoutsos Isidore, Huynh Tien, Miranda Kevin, Tsirigos Aristotelis, McHardy Alice, Platt Daniel. Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes. Proc Natl Acad Sci 2006;103(17):6605–10.

[7] Gaziev AI, Shakhaev GO. Nuclear mitochondrial pseudogenes. Mol Biol 2010;44(3):405–17.

[8] Hirotsune S, Yoshida N, Chen A, et al. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature 2003;423(6935):91–6.

[9] Tam OH, Aravin AA, Stein P, et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 2008;453(7194):534–8.

[10] Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010;465(7301):1033–8.

[11] Betrán E, Wang W, Jin L, Long M. Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. Mol Biol Evol 2002;19(5):654–63.

[12] van Baren MJ, Brent MR. Iterative gene prediction and pseudogene removal improves genome annotation. Genome Res 2006;16(5):678–85.

[13] Rodier Francis, Campisi Judith. Four faces of cellular senescence. J Cell Boil 2011;192(4):547–56.

[14] Castelo-Brancol Pedro, Zhang Cindy, Lipman Tatiana, et al. Neural tumor-initiating cells have distinct telomere maintenance and can be safely targeted for telomerase inhibition. Clin Cancer Res 2011;17:111–21.

[15] Günesand Cagatay, Rudolph Lenhard. The role of telomeres in stem cells and cancer. Cell 2013;152(3):390–3.

[16] Hirashima Kyotaro, Migita Toshiro, Sato Shigeo, Muramatsu Yukiko, Ishikawa Yuichi, Seimiya Hiroyuki. Telomere length influences cancer cell differentiation in vivo. Mol Cell Biol 2013. http://dx.doi.org/10.1128/MCB.00136-13.

[17] Gomes NM, Ryder OA, Houck ML, et al. Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. Aging Cell 2011;10(5):761–8.

[18] Hsu LY, Benn PA, Tannenbaum HL, Perlis TE, Carlson AD. Chromosomal polymorphisms of 1, 9, 16, and Y in 4 major ethnic groups: a large prenatal study. Am J Med Genet 1987;26(1):95–101.

[19] Lowry David, Willis John. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLOS Biol 2012;10(1):10.

[20] Heidari Nastaran, Phanstie Douglas, He Caho, Grubert Fabian, Jahanbanian Fereshteh, Kasowski Maya, et al. Genome-wide map of regulatory interactions in the human genome. Genome Res 2014. http://dx.doi.org/10.1101/gr.176586.114.

[21] Milstein C. The hybridoma revolution: an offshoot of basic research. BioEssays 1999;21(11):966–73.

[22] Griffiths Anthony, Gelbart William, Miller Jeffrey, Lewontin Richard. RFLP mapping. In: Modern genetic analysis. New York: W. H. Freeman; 1999 [Chapter 11].

[23] Cox David, Burmeister Margit, Price Roydon, Kim Suwon, Myers Richard. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 1990;250:245–50.

[24] Segal G, Liu B, Vega JM, Abbo S, Rodova M, Feldman M. Identification of a chromosome-specific probe that maps within the Ph1 deletions in common and durum wheat. Theor Appl Genet 1997;94(8):968–70.

[25] Omoto Charlotte, Lurquin Paul F. Linkage and mapping: gene discovery. In: Genes and DNA: a beginner's guide to genetics and its applications. New York Chichester, West Sussex: Columbia University Press; 2004 [Chapter 9].

[26] Pasternak Jack J. The genetic system: Mendel's laws of inheritance and genetic linkage. An introduction to human molecular genetics – Mechanisms of inherited diseases. Wiley-Liss; 2005 [Chapter 3].

Additional resources

[1] Alberts B et al. Molecular biology of the cell. 4th ed. Hamden, CT, USA: Garland Science; 2002.

[2] Anthony JF, Wessler Susan R, Lewontin Richard C, Carroll Sean B. Introduction to genetic analysis. Gordonsville, VA, USA: W.H. Freeman & company; 2007.

[3] Pierce Benjamin A. Genetics: a conceptual approach. 2nd ed. Gordonsville, VA, USA: W.H. Freeman & company; 2005.

[4] Omoto Charlotte, Lurquin Paul F. Genes and DNA: a beginner's guide to genetics and its applications. New York, Chichester, West Sussex: Columbia University Press; 2004.

[5] Eberhard Passarge. Color atlas of genetics. 3rd ed. Stuttgart, New York: Thieme; 2007.

[6] Rédei GP. Encyclopedia of genetics, genomics, proteomics and informatics. 3rd ed. New York, USA: Springer; 2008.