31

The Future of Human Cytogenetics

Continued improvements in cytogenetic technique have consistently revealed chromosome aberrations not detectable by earlier methods. The likeliest repository of additional aberrations in human populations remains the same as before: spontaneous abortions, stillbirths, infants with multiple malformations, mentally retarded individuals, and cancers, examined using still newer methods. For example, mental retardation affects up to 3% of the population. Detectable chromosome imbalance accounts for a large percentage of severe mental retardation (IQ <55) and perhaps 5–10% of milder retardation (IQ 55–70), but the cause is unknown in many cases. Flint et al. (1995) have taken a novel approach, based on evidence that the subtelomeric regions are generally prone to recombination and that methods now exist for detecting unequal recombinational events by probing the DNA of flow-sorted chromosomes with polymorphic markers. They estimate that 6% of unexplained mental retardation

O. J. Miller et al., *Human Chromosomes*

may be due to small chromosome changes in the subtelomeric regions alone. Better methods are also needed to identify paracentric insertions, which carry a 15% risk of duplications or deletions arising during meiotic segregation in carriers. Currently, as many as 90% of these insertions may be undetectable by present methods, based on their frequency in the two chromosomal regions where they can be most readily detected.

Unsolved Problems

Some of the most fundamental questions in human cytogenetics have yet to be answered. The role of telomere shortening in cell aging and carcinogenesis has become clear, but is telomere shortening also a major factor in other diseases of advancing age, such as non-insulin-dependent diabetes, atherosclerosis, and hypertension, and if so, how can its harmful effects be circumvented? What role do abnormalities of telomeres, centromeres, and the spindle play in nondisjunction? What are the causes of nondisjunction, and what is responsible for the profound maternal age effect?

How do trisomies, duplications, and deletions produce their phenotypic effects? That is, what are the critical genes responsible for these gene dosage effects? Cytogeneticists have approached the problem by trying to define the critical region for Down syndrome and for various duplication or deletion syndromes, and to use positional cloning to identify, in each critical region, specific genes with a major dosage effect. This could provide clues to the metabolic, signaling, or other pathways involved in the developmental errors. This exciting approach has already yielded insights into several deletion and imprinting syndromes and holds much promise. One should not forget that some genes act by suppressing other genes. The increased gene dosage associated with trisomy for a chromosome may exert its major effects by inhibiting genes elsewhere in the genome. An exciting example of this is the recent study of the synchrony of replication of the two alleles at four gene loci on three different autosomes (*MYC* on 8, *RB1* on 13, *TP53* and *HER2* on 17) in cultured amniotic fluid cells. In normal cells, the alleles at each of these four loci replicated synchronously, but in 21 trisomic cells they replicated asynchronously (Amiel et al., 1998). The presence of an earlier- and a later-replicating allele suggests that one allele has been inactivated. A reduction in gene dosage at these loci could be responsible for the increased cancer risk in Down syndrome. How many other genes are similarly affected in trisomy 21, and what features of Down syndrome are they responsi-

ble for? If trisomies or partial trisomies of other chromosomes also lead to replication asynchrony of particular genes at distant loci and reduced dosage of their products, our views on how chromosome imbalances lead to phenotypic abnormality could be profoundly changed.

What triggers late replication of both alleles of tissue-specific genes and of one allele of imprinted genes? What triggers X inactivation, and how do genes on the X chromosome or an attached autosomal segment escape inactivation? How are changes in the packaging of chromosomes brought about during cell differentiation, and how is this particular chromatin state maintained through mitosis? What are the evolutionary and mechanistic relationships between imprinting and X inactivation? The inactivation of most genes on one of the two X chromosomes in females equalizes the level of expression of these genes in males and females. A different dosage compensation mechanism is necessary to equalize the expression levels of genes on the single active X in both sexes and on the two expressed copies of most autosomal genes, but virtually nothing is known about this process. An attractive idea is that dosage compensation between X-linked and autosomal genes involves a doubling of the expression level of genes on the X (Ohno, 1967). Graves et al. (1998) have reviewed this topic and presented impressive evidence. The level of expression of the *Clc4* gene is twice as high in *Mus spretus*, in which it is on the X chromosome, as it is in *Mus musculus*, in which it is on an autosome. It is intriguing that dosage compensation between XY and XX in *Drosophila* is mediated the same way, suggesting that this very poorly understood autosome: X dosage compensation mechanism is far more highly conserved than anyone realized, and providing a lead for future research in humans.

Genome Organization

Some genes with closely related functions, which probably arose by duplication events, have maintained their contiguous locations throughout their very long evolutionary histories. A long-standing question in cytogenetics is whether the map position of any gene simply reflects its evolutionary history or whether there are functional constraints that maintain some tightly linked groups of genes. In general, the former seems to be the case, even for genes of similar function, such as cyclins or cyclin-dependent kinases (Chapter 2). The more than 80 genes encoding the structural proteins of the ribosome are widely scattered throughout the genome, even though they are coordinately regulated (Feo et al., 1992).

Table 31.1. Gene Clusters Maintained by Functional Constraints

Gene cluster	Symbol	Size (kb)	Location	Type of band
Hemoglobin α-globin	*HBBA*	70	16p13.3–pter	R (T)
Hemoglobin β-globin	*HBBC*	65	11p12	G
Homeobox A	*HOXA*	>100	7p15.3	R
Homeobox B	*HOXB*	>100	17q21.3	R
Homeobox C	*HBAC*	>60	12q13.3	R
Homeobox D	*HOXD*	>80	2q31	R
Major histocompatibility	*MHC*	>2000	6p21.3	R
Immunoglobulin κ light chain	*IGK*	160	2p11.2–p12	—
Immunoglobulin λ light chain	*IGL*	140	22q11.12	R
Immunoglobulin heavy chain	*IGH*	1200	14q32.33	R (T)
T-cell receptors α and δ	*TCRA/D*	140	14q11–q12	—
T-cell receptor β	*TCRB*	840	7q32–q33	—
T-cell receptor γ	*TCRG*	120	7p15	R
Odorant receptor family on 7		>100	7p22	R (T)
Odorant receptor family on 16		>100	16p13	R (T)
Odorant receptor family on 17		>100	17p13	R (T)
Imprinted cluster on 11			11p15.5	
Imprinted cluster on 15			15q11–q13	

However, there are several striking examples in which the position of a gene within a cluster is highly correlated with temporal and tissue-specific patterns of expression (Table 31.1). The best known may be the α- and β-globin gene clusters on chromosomes 16 and 11, respectively.

The genes at the 5′ end of each globin cluster are expressed earlier in development, and the genes at the 3′ end are expressed later. Each gene has its own promoter, but each cluster is under the control of a *locus control region* (LCR) that determines its chromatin structure, time of replication, and expression pattern. Tanimoto et al. (1999) used the powerful Cre-loxP recombination technique to invert either the human β-globin cluster of five genes or the LCR itself (Fig. 31.1). The ε-globin gene, normally at the 5′ end (nearest the LCR), is no longer expressed during the yolk-sac stage of erythropoiesis when it is separated from the LCR, and LCR activity itself is markedly diminished if the LCR is inverted. A similar mechanism may underlie the maintenance of small gene clusters in imprinted regions: Upstream regulatory elements must be maintained in addi-

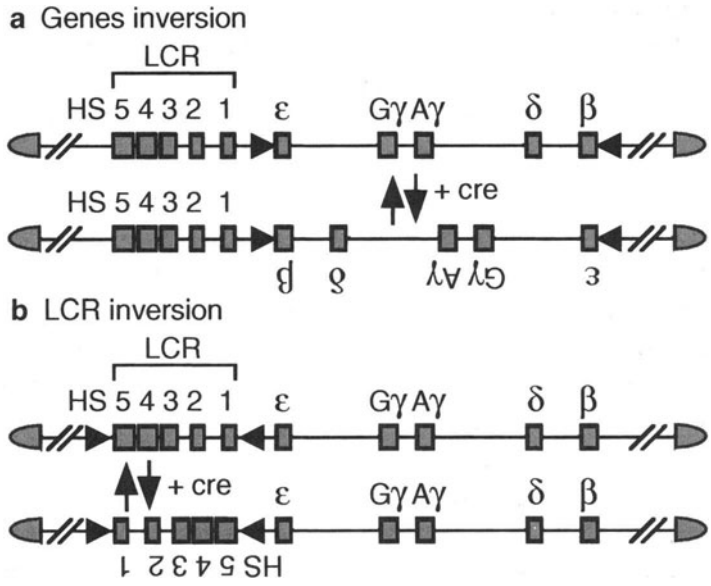

Figure 31.1. Gene manipulation: inversion of either the five-gene β-globin gene cluster (a) or the locus control region (LCR) (b), used to demonstrate the importance of order within each of these regions for gene function (reprinted with permission from Nature 398, p 345, Tanimoto et al., copyright 1999, Macmillan Magazines Limited).

tion to the genes themselves (Chapters 21 and 22). Homeobox (*HOX*) genes play critical roles in early embryogenesis. The four *HOX* gene clusters, each with about 8–11 contiguous *HOX* genes, show a very tight temporal and spatial correlation between gene order and function; that is, the genes are activated sequentially, from one end of each array to the other. Fine mapping of the four clusters places them all in R-bands (Apiou et al., 1996).

A different functional requirement has maintained the very large immunoglobulin heavy- and light-chain gene clusters and the T-cell receptor gene clusters. Here, site-specific breakage, elimination of much DNA, and V(D)J rejoining produce the exceptionally large number of antibody specificities and histocompatibility antigen cell surface receptor specificities so important for our resistance to infectious disease and cancer (Tonegawa, 1983). How many more constraints on genome organization will be found? Will their study reveal new genetic mechanisms that are as unexpected as imprinting?

A candidate for just such a discovery is the olfactory (odorant) receptor gene family, first identified in 1991 in the rat by Buck and Axel and then in 1992 in

humans by several groups. There are several hundred copies of these genes. Nevertheless, there appears to be tight control over the entire family, because each olfactory neuron expresses very few olfactory receptor genes and only a single allele of each (Chapter 22). This is the basis of our ability to distinguish thousands of different odors. This complex regulation might involve imprinting the entire gene family in both oogenesis and spermatogenesis and activation of a single allele at a single locus in each olfactory neuron, but how is the activation of a different olfactory receptor gene in each cell achieved? An explanation would be at hand if all the family members were in a single long tandem array. However, most of the potentially functional copies are in tandem arrays hundreds of kilobase pairs long on chromosomes 7, 16, and 17. Additional smaller clusters are present at more than 20 locations throughout the genome, most of them in terminal bands (Rouquier et al., 1998). This may be related to another peculiar feature of this gene family, the presence of a single open reading frame (exon) in each olfactory receptor gene. No intron divides the exon, which is about 1 kb long. Intronless genes usually arise from genes with introns by reverse transcription of a processed mRNA and random reintegration into the genome. Such an origin might account for the widespread distribution of the gene family, but the precise regulation of these genes remains a mystery.

Genetic recombination is suppressed in a few regions of the genome. Genes in these regions are said to show *linkage disequilibrium*. One of these regions is on 17q21. It contains the *RNU2* array of 6–30 copies of the *U2* small nuclear RNA genes and the adjacent 175-kb region containing the tandemly duplicated *NBR1*, *LBRCA1* pseudogene, *NBR2*, and *BRCA1* genes. This 200 to 400-kb region shows complete suppression of crossing over, and only two major haplotypes (different allelic forms) of this region were observed in 275 Europeans and 34 Asians, indicating that crossover suppression in this region has persisted for at least the last 100,000 years (Liu and Barker, 1999). This region is rather close to the *HOXb* locus, but no functional relationship between the two clusters or their behavior is known.

Methods are needed to enable recognition of functionally important gene clusters. Close similarity in sequence (reflecting a common origin) will be readily identifiable, once the Human Genome Project is completed and the complete nucleotide sequence of every chromosome is known. However, genes of quite different sequences may remain closely linked because of shared regulatory controls, such as a promoter or LCR. Working out additional features of genome organization that delineate functional domains is a task for the future.

Mapping the human genome is still far from complete, and this is particularly so for locating and identifying all the transcribed sequences (genes). Yeast artificial chromosomes (YACs) may become a powerful tool for analyzing closely linked genes or gene clusters. All the human genes in a YAC are transcribed in yeast cells, which lack the tissue-specific regulatory systems that suppress the expression of most genes in human cells (Still et al., 1997). This technique provides a unique approach to identifying all the genes in a particular region and may be generally useful in completing our genetic map.

The highly nonrandom organization of the human genome presents the cytogeneticist with a bewildering array of unanswered questions. How did the 17-fold greater density of genes in the most GC-rich chromosome bands arise, how is it maintained, and what is its functional significance? Why are genes concentrated in subtelomeric locations, and is that concentration related to the high frequency of crossing over in these regions or to the fact that minisatellites almost always originate in these regions? The most gene-rich region yet defined is less than 50 kb long, in band 9q34.2. This region, called the *surfeit* locus, contains six housekeeping genes that have no sequence similarity or functional similarity. Five of the genes are separated from each other by tiny spacer sequences 97–302 bp long. Even the gene order is conserved in mammals and birds (Duhig et al., 1998), although there is no evidence that this particular gene cluster is maintained by functional constraints such as those involved in the globin, immunoglobulin, T-cell receptor, and homeobox genes. However, tighter clustering has been achieved by sharing CpG islands, so that four rather than six of these are sufficient. That is, there is a CpG island at the 5′ end of each gene; the middle two CpG islands are each shared by two genes. This has the interesting consequence that the two genes sharing one CpG island are then transcribed in opposite directions (from opposite strands). Does such sharing have implications for gene regulation, and is there a preferred (nonrandom) orientation of transcribed strands along a chromosome?

Directions

The enormous successes of molecular cytogenetics in the last decade provide a clear indication of a number of developments to be expected in the future. Contigs based on ordered, overlapping cloned DNA fragments propagated in cosmids and bacterial and yeast artificial chromosomes will facilitate rapid mapping of new disease genes and their identification by positional cloning,

aided by libraries of ESTs (expressed sequence tags) arrayed on DNA chips. Normalized cDNA libraries, containing nearly equal representation of all expressed sequences from sequential developmental stages, will speed understanding of the temporal and tissue specificity of many genes and perhaps throw light on the functional significance of chromosome organization. There will be a steady increase in the number of painting and other probes for specific chromosome segments, bands, and the breakpoint regions of specific rearrangements that cause cancer.

The introduction of human genes, DNA segments, or artificial chromosomes into transgenic or transgenomic mice may permit in vivo analysis of the effects of certain trisomies, duplications, and deletions on differentiation, providing needed animal models. Portions of various human chromosomes have been introduced into mouse embryonic stem cells by microcell-mediated cell fusion/chromosome transfer. Viable chimeric mice have been produced from these, and some of the transferred genes are expressed in their correct tissue-specific way in the adult tissues (Tomizuka et al., 1997). Methods are being developed for the conditional silencing of target genes. Unlike the constitutive silencing produced by standard gene targeting (gene knockout) techniques, conditional silencing methods will permit a gene to be turned off at a specific time and in a specific cell type (Porter, 1998). This could permit the study of otherwise lethal genes and those producing severe birth defects, as in various trisomies or deletions.

New Technology

Improvements in microscopes, development of *charge-coupled device* (CCD) cameras and computer-controlled digital image capture of very faint fluorescent signals, and use of computer interfaces that permit optical sectioning and three-dimensional reconstruction of nuclei have greatly expanded the usefulness of the cytogeneticist's classic instrument, the light microscope. This has been possible only because of equally impressive developments in the preparation of molecular probes, the DNA fragments essential for in situ hybridization. Continued developments in these areas are expected, such as an expansion of interphase and preimplantation cytogenetics.

More revolutionary advances are on the horizon, with some already coming into use. DNA chip technology is one of these. Using the techniques developed for making silicon microchips for computers, it is possible to prepare many

copies of orderly arrays of tens of thousands of DNA fragments or oligonucleotides on a small nitrocellulose filter, glass slide, or silicon chip. The power of this approach is illustrated by the discovery of hundreds of cell cycle–regulated genes in yeast (Spellman et al., 1998). There are now 800 known, in contrast to 104 in 1997. Many, if not most, of these genes that are required only once per cell cycle will have human homologues that can now be rather quickly identified.

The same chip technology has been used to make microelectrophoresis channels that can be filled with a replaceable gel and used to type short tandem repeat genetic markers in just 30 seconds (Schmalzing et al., 1997). This could greatly speed the mapping of translocation breakpoints, the identification of new disease genes, and the understanding of multifactorial disorders like diabetes and hypertension. Chip arrays can be used to screen quickly and cheaply for specific rearrangements that cause cancer (Hacia et al., 1998). A microarray of 10,000 cDNAs could analyze the expression of thousands of genes in a single experiment, creating a need for new types of data management and analysis (Ermolaeva et al., 1998).

The Human Genome Project and its rapidly growing body of users are creating ever more massive databases. Their use has been greatly facilitated by ready Internet access. Frequent reference to these databases will become an essential part of keeping abreast of relevant advances, for both research and clinical cytogenetics laboratories. Bibliographic information can be obtained from the National Library of Medicine: PUBMED at http://www.ncbi.nlm.nih.gov. This is perhaps most easily accessed through the NCBI Entrez Browser at http://www.ncbi.nlm.nih.gov/Entrez/. The Online Mendelian Inheritance in Man (OMIM) Home Page, at http://www.ncbi.nlm.nih.gov/Omim, is a continuously updated catalogue of human genes and genetic disorders. The Genome Data Base (GDB) at http://www.gdb.org/ has molecular and mapping data.

Advances in cell culture will include discovery of mitogens for additional types of differentiated G0 cells, enabling them to enter mitosis. Fibroblast growth factor 4 (FGF4) has recently been shown to permit the development of permanent placental trophoblast stem cell lines that can differentiate into trophoblast subtypes (Tanaka et al., 1998). The introduction of genes for positive and negative selection of specific chromosomes or chromosome segments should make it possible to develop cell lines with specific deletions or monosomies. These could be used to study altered gene dosage or regulation, to show which aneuploidies are not viable at a cellular level, and to learn the reason. An immortal testicular Sertoli cell line has been produced in the mouse that supports the

differentiation of meiotic and even postmeiotic cells from premeiotic germline cells (Rassoulzadegan et al., 1993). The ability to induce human germline cells to follow this meiotic pathway would provide a powerful impetus to the analysis of causes of nondisjunction and the mechanism of imprinting. Human pluripotent stem cell lines have recently been developed (Thomsen et al., 1998). Their potential use in basic research and cell or gene therapy will depend upon significant input from cytogeneticists.

Human artificial chromosomes have recently been created. Mitotically stable minichromosomes only 4–9 Mb long have been generated from the Y chromosome by telomere-directed chromosome breakage (Heller et al., 1996). Completely artificial chromosomes of comparable length have been produced from telomeric DNA, genomic DNA, and synthetic α-satellite arrays from chromosomes 17 and Y or chromosome 21; these are mitotically and cytogenetically stable (Harrington et al., 1997; Ikeno et al., 1998). Such artificial chromosomes may become useful for long-term correction of genetic defects, by serving as vehicles for the stable introduction of a functional copy of the missing gene. They could also be very useful research tools for investigating causes of nondisjunction in cell culture systems. A microchromosome that arose in a man with the CREST form of scleroderma was transmitted to a normal daughter, indicating meiotic as well as mitotic stability for such chromosomes (Haaf et al., 1992).

References

Amiel A, Avivi L, Gaber E, et al. (1998) Asynchronous replication of allelic loci in Down syndrome. Eur J Hum Genet 6:359–364

Apiou F, Flagiello D, Cillo C, et al. (1996) Fine mapping of human HOX gene clusters. Cytogenet Cell Genet 73:114–115

Duhig T, Ruhrberg C, Mor O, et al. (1998) The human *Surfeit* locus. Genomics 52:72–78

Ermolaeva O, Rastogi M, Pruitt KD, et al. (1998) Data management and analysis for gene expression arrays. Nat Genet 20:19–23

Feo S, Davies B, Fried M (1992) The mapping of seven intron-containing ribosomal protein genes shows they are unlinked in the human genome. Genomics 13:201–207

Flint J, Wilkie AOM, Buckle VJ, et al. (1995) The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. Nat Genet 9:132–139

Graves JAM, Disteche CM, Toder R (1998) Gene dosage in the evolution and function of mammalian sex chromosomes. Cytogenet Cell Genet 80: 94–103

Haaf T, Sumner AT, Köhler J, et al. (1992) A microchromosome derived from chromosome 11 in a patient with the CREST syndrome of scleroderma. Cytogenet Cell Genet 60:12–17

Hacia JG, Makalowski W, Edgemon K, et al. (1998) Evolutionary sequence comparisons using high-density oligonucleotide arrays. Nat Genet 18:155–158

Harrington JJ, Bokkelen G Van, Mays RW, et al. (1997) Formation of de novo centromeres and construction of first-generation human artificial chromosomes. Nat Genet 15:345–355

Heller R, Brown KE, Burgtorf C, et al. (1996) Mini-chromosomes derived from the human Y chromosome by telomere directed chromosome breakage. Proc Natl Acad Sci USA 93:7125–7130

Ikeno M, Grimes B, Okazaki T, et al. (1998) Construction of YAC-based mammalian artificial chromosomes. Nat Biotechnol 16:431–439

Liu X, Barker DF (1999) Evidence for effective suppression of recombination in the chromosome 17q21 segment spanning *RNU2–BRCA1*. Am J Hum Genet 64:1427–1439

Ohno S (1967) Sex chromosomes and sex-linked genes. Springer-Verlag, New York

Porter A (1998) Controlling your losses: conditional gene silencing in mammals. Trends Genet 14:73–79

Rassoulzadegan M, Paquis-Fluckinger V, Bertino B, et al. (1993) Transmeiotic differentiation of male germ cells in culture. Cell 75:997–1006

Rouquier S, Taviaux S, Trask BJ, et al. (1998) Distribution of olfactory receptor genes in the human genome. Nat Genet 18:243–250

Schmalzing D, Koutny L, Adoutian A, et al. (1997) DNA typing in thirty seconds with a microfabricated device. Proc Natl Acad Sci USA 94: 10273–10278

Spellman PT, Sherlock G, Zhang MO, et al. (1998) Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 9:3273–3297

Still IH, Vince P, Cowell JK (1997) Direct isolation of human transcribed sequences from yeast artificial chromosomes through the application of RNA fingerprinting. Proc Natl Acad Sci USA 94:10373–10378

Tanaka S, Kunath T, Hadjantonakis A-K, et al. (1998) Promotion of trophoblast stem cell proliferation by FGF4. Science 282:2072–2075

Tanimoto K, Liu Q, Bungert J, et al. (1999) Effects of altered gene order or orientation of the locus control region on human β-globin gene expression in mice. Nature 398:344–348

Thomsen JA, Itskowitz-Eldor J, Shapiro SS, et al. (1998) Embryonic stem cell lines derived from human blastocysts. Science 282:1145–1147

Tomizuka K, Yoshida H, Uejima H, et al. (1997) Functional expression and germline transmission of a human chromosome fragment in chimaeric mice. Nat Genet 16:133–143

Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302: 575–581