REVIEW ARTICLES

The Genetics of Childhood Disease and Development: A Series of Review Articles

The following article is the first in our new series "The Genetics of Childhood Disease and Development." It provides an introduction to the rapidly expanding field of genomics in pediatric research and the existing and potential contribution of that field to the understanding, diagnosis, and treatment of childhood disorders.

Alvin Zipursky
Editor-in-Chief

Genomics and Pediatric Research

ANDREW P. BORIGHT, JUHA KERE, AND STEPHEN W. SCHERER

Department of Medicine, University Health Network, University of Toronto [A.P.B.], and Department of Genetics and Genomic Biology, The Hospital for Sick Children, and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada [S.W.S.]; and Department of Biosciences at Novum and Clinical Research Centre, Karolinska Institute, Stockholm, Sweden [J.K.]

In the past few years impressive amounts of DNA sequence have been generated from the genomes of many species. It is now safe to claim that a "finished" version of the human genome will be available by mid-2003, just 50 y after the structure of DNA was first described (a "draft" or fragmented version of the human genome sequence has been available since 2001). There has also been much hyperbole surrounding announcements that described these achievements, and their significance for biomedical research and the human condition. Through this commentary we will introduce the series "Genetics of Childhood Disease and Development" to the pediatric clinician and scientist and provide a review on the more exciting observations of the Human Genome Project (HGP). We also aim to temper some earlier statements that need further clarification, and to make predictions of where this new science might have its greatest impact. For additional reference, we have highlighted the relevant original or review articles on these topics. A thorough overview of

facts, figures, slides, educational and media presentations, and the most comprehensive set of genome-related links can be found at the "Human Genome Project Information" site (<http://www.ornl.gov/hgmis/>).

Genomics is the study of the structure and function of an organism's genetic material (the genome) including DNA, genes, and chromosomes (Fig. 1). As implied by its definition, genome science is a large-scale undertaking, being highly dependent on technology and automation. Genomic research can be considered largely exploratory, as it does not need to follow the hypothesis-driven rules of more classic biomedical research. In fact, for genomic research to be effective it need only generate information (or technological innovations) from which new hypotheses or data sets can be derived. During the past decade, this new philosophy for performing biologic research has become generally, albeit often hesitantly, accepted, and genomic information now sets the foundation for many areas of scientific investigation.

For the 12-mo period after the February 2001 publications describing a draft version of the human genome sequence (1, 2), we determined, using a liberal definition, that genomic information or technologies had been used in no fewer than 54 of 240 (23%) of articles published in *Pediatric Research*. Although this statistic is relatively high, in particular for a journal known historically to emphasize research in pediatric physiology, we anticipate that this number will continue to increase as the understanding of genomic information and

Received July 12, 2002; accepted July 24, 2002.

Correspondence: Stephen W. Scherer, Ph.D., Department of Genetics, Room 9107, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, M5G 1X8, Canada; e-mail: steve@genet.sickkids.on.ca

Supported by grants from Canadian Institutes of Health Research (CIHR), Genome Canada, and the Howard Hughes Medical Institute. Stephen W. Scherer is an Investigator of the Canadian Institutes of Health Research and International Scholar of the Howard Hughes Medical Institute.

DOI: 10.1203/01.PDR.0000042461.81761.0B

Figure 1. Study of the variation and mutation in genes associated with phenotype underlies traditional genetic research. Genome, transcriptome, and proteome variation also influence phenotype, and advances in new technologies allow large amounts of biologic information to be collected for disease studies. Definitions of different genome-era terms are given.

availability of reagents continues to infiltrate the larger scientific community. In saying this, as reviewers of grants and manuscripts in both the genomics and pediatric research fields, it is surprising to see how often genomic information is still not being used to the fullest extent. The types of information and molecular reagents used for specific studies tend to be drawn from some of the more general observations of the initial analyses of the human genome including the following areas.

NUMBER AND DISTRIBUTION OF GENES

For experimental biology and molecular diagnostics the most important part of what constitutes genomes is the gene. It is estimated that less than 3% of the human genome encodes protein-coding genes, of which there seem to be roughly 30,000–40,000 (1, 2). The precise number continues to be hotly debated (3). It could very well be that there are upward of 50,000 or more genes in the genome, or perhaps as few as 28,000. Some of the discrepancies in numbers can be attributed to different interpretations of the definitions being used. To provide a basis for the estimates, different data sets of partial and complete sequences as well as gene models based on computer predictions are often used. Moreover, these data sets are still largely incomplete. For example, at present we know the complete sequence from true start to finish for less than 50% of the 30,000–40,000 genes predicted using computer algorithms. Even less is known about the cellular function of their protein products. In the characterized human collection the most is known about those 1,000 or so genes that have been characterized with respect to human disease (see <http://www.ncbi.nlm.nih.gov/Omim/> and <http://www.geneclinics.org/>).

Whatever the final number of genes turns out to be, it will likely be lower than some might think our species might deserve. One of the most interesting observations to come

out of genome sequencing is that the gene number does not necessarily correlate with organismal complexity or evolutionary hierarchy. For example, the fruit fly *Drosophila melanogaster* has 13,000 genes and comprises 10 times more cells than the worm *Caenorhabditis elegans*, which has 19,000 genes, and the simple plant *Arabidopsis thaliana* has approximately 26,000 genes, near equivalent to the human. To the reductionist, *Homo sapiens* might be seen as no more than a fruit fly plus a worm, or perhaps, a plant supplemented with a few extra cerebral genes. However, humans seem to have a greater combinatorial diversity of proteins than other species, which arises, in part, as a result of alternative spliced forms of upward of 40% of the defined gene set (4). Therefore, in some cases, the classic “one gene, one protein” dogma is being redefined. It is also interesting that genes need not be uniformly distributed along chromosomes. In the human genome, chromosome 19 seems to be the most gene-rich. Chromosomes 18 and 21 contain the fewest genes, which might have been predicted inasmuch as trisomy of either of these chromosomes alone can result in fetal viability. A major focus of continued research of genome projects is to properly annotate and isolate all of the genes and their regulatory regions. The availability of these resources is vital for further biomedical investigation, including gene therapy, functional experimentation, and microarray expression studies, as well as protein production for structural analysis and therapeutics.

REPETITIVE DNA AND GENOME DUPLICATION

It has been known for many years that the human genome is riddled with repetitive DNA elements. Analysis of the human genome now amply indicates that up to 45% of the sequence is found in the form of long and short interspersed repeats, retroviral repeats, transposable elements, satellite sequences in heterochromatin regions, and dynamic trinucleotide repeats (often involved in diseases demonstrating genetic anticipation). The role of all of these classes of DNA elements is still mostly enigmatic, and for lack of a better term the nongenic DNA is often dubbed “selfish” DNA (DNA whose purpose seems related only to self-replication) or “junk” DNA. However, with the development of new genomic tools these elements have been implicated in mutagenic processes (5–7). Besides providing for a better understanding of the organization of these repeats, perhaps the most surprising finding of genome analysis is the discovery that upward of an additional 5% of the human genome sequence has arisen through recently occurring large-scale duplication. Many of these events occurred only since primates evolved and in some cases they are specific to the human genome. The phenomena of genomic duplication is observed in two forms: one mediated at the DNA level (segmental duplication, which can encompass hundreds of kilobases of DNA) and the other at the RNA level (retrotransposition or copying of RNA to DNA followed by reinsertion into the genome). Each mechanism can yield related (or paralogous genes) in more than one copy in the genome, which can sometimes be involved in the same disease. A noteworthy example is the finding that related genes on chromosomes 1 and 14 of the presenilin family

cause Alzheimer's disease (8). The role of intrachromosomal genomic segmental duplications in microdeletion syndromes is now well established (5, 9). The pediatric syndromes of DiGeorge, Williams-Beuren, Smith-Magenis, Prader-Willi, Angelman, and others (5) all arise because of deletions of millions of nucleotides of DNA of defined chromosomal regions (and the intervening genes) bounded by segmental duplications. Chromosomal rearrangements can arise because misalignment of the near-identical repeated sequences can occur during pairing at meiosis and mitosis. The segment of DNA between the repeated elements does not have a proper partner and is then prone to being deleted, duplicated, or inverted. Paradoxically, although such duplicated regions of the genome should be considered suspect in any disease gene search, they are often underrepresented or not represented in the current draft versions of the genome owing to difficulties in their characterization because of near-identical DNA sequences over large chromosomal regions (10, 11). Nevertheless, the existing genomic maps have recently shown that segmental duplication-mediated chromosomal deletions account for Sotos syndrome (12) and forms of male infertility (13), and duplications and inversions exist in panic and anxiety disorder (14) and Williams-Beuren syndrome (15), respectively. It is safe to predict that segmental duplications also will be culprit to other congenital and somatic diseases, as well as underlie apparently nonpathogenic genomic variations, as has been found at an olfactory receptor locus on chromosome 8 (16, 17). Many new genomic resources and technologies, including DNA microarray analyses, will certainly expedite discovery of unbalanced chromosome rearrangements for prenatal, perinatal, and carrier screening in the diagnostic setting (18, 19).

DNA SEQUENCE VARIATION

The HGP has unveiled many types of variation in our 3.2 billion-bp genetic code with single nucleotide changes (or SNPs) as the most frequent, occurring one in 1000 nucleotides (1, 20). This translates into the apparent finding that each of us in the world regardless of race or geography has 99.9% DNA sequence identity. Therefore, on average, each genome contains approximately 3 million SNPs. Other DNA sequence variations include copy number changes, insertions, deletions, duplications, translocations, inversions, and more complex rearrangements. These genomic changes do not occur as frequently as SNPs, but the determination of their true frequency in the general population and in disease states will now be possible with the new genomic technologies available. Of course, genetic variation is the basis of evolution, and it can contribute to health, unique traits, or disease. Pediatricians are also aware that their patients are predisposed to different diseases, metabolize drugs differently, and vary regarding dose-response relationships for pharmaceuticals that might have toxicologic implications (21). In an often-cited paper, it has been reported that 1.8 million people were admitted to the hospital for adverse drug events in 1994, in which more than 100,000 deaths

resulted (22). The reasons why a patient may or may not respond to a drug or suffer an adverse reaction are many, including misdosing, drug-drug interaction, drug allergy, and medication error. The technological advances, availability of genetic markers, and higher resolution SNP maps generated by the HGP will now provide a more precise estimation and understanding of a patient's individual genetic predisposition for an adverse drug response (23, 24). Other biomedical applications will include identifying variations in immune response genes, which will have implications for transplantation and vaccine development. To assist in DNA sequence variation and disease gene studies, during the past decade the HGP has already generated many high-resolution genetic maps of the human genome (20, 25, 26). Moreover, there is currently an initiative that will elucidate patterns of blocks of SNPs along chromosomes (called haplotypes) in different populations, leading to the generation of catalogs of the genetic variation most important to health and disease in a target group being studied. This latest stage of the HGP has been dubbed the haplotype map or "HapMap" (27). Having an organized SNP haplotype map should increase the statistical power available for making an association between a gene and a complex disease phenotype. However, even with such promise the project has not been without controversy (28) because of concerns about the true biomedical value, the cost associated with completing it, and competition with the private sector (29), as well as the ethical concerns associated with determining patterns of genetic variation across populations (28).

DNA SEQUENCE CONSERVATION AND COMPARATIVE GENOMICS

Data from the genome projects have demonstrated that an extremely powerful approach to finding genes is through comparison of the DNA sequence of two organisms to search for shared elements (30, 31). If a segment of DNA has been conserved throughout evolution it almost always encodes a gene or a regulatory element controlling the expression of that gene. Almost every known human gene can also be found in the genomes of other mammalian species (*e.g.* mice and dogs), with the similarity of the two sequences often exceeding 85%. The genomes of human and chimpanzee (including noncoding sequences) are estimated to be near to 99% identical (32, 33). Comparative genomic studies have also revealed that some genes, including those that encode proteins involved in replicating DNA (and are thus absolutely crucial to survival of every organism), are highly conserved in every species from human all the way down the evolutionary tree to yeast and bacteria. Just as interesting, comparative analyses of bacterial genomes have revealed that prokaryotes share genes among themselves quite promiscuously (on an evolutionary time scale) (34). They can behave, in some ways, like a single ancient and global species. However, contrary to what was predicted in the public HGP manuscript (2) and overly publicized in the popular press, the direct vertical transfer of genes from bacteria to human is not commonplace (35).

A combination of bioinformatics analysis and genetic engineering has shown that the minimal number of nonessential and essential genes necessary for life in *Mycoplasma genitalium*, the simplest known cell, is between 265 and 350 (36). Using similar comparative genomics strategies, biologically relevant, strain-specific pathogen and antibody resistance genes can be readily identified (37, 38). Continuing the work of 19th-century natural biologists, who primarily cataloged species, genome scientists will now sequence those genomes to provide a wealth of data for further investigation. The ability for refinement of predictions of gene and protein function increases with the amount of information available for comparative studies. There are presently 80 genomes completed or nearly completed, with most being of prokaryotic origin (see <http://www.ncbi.nlm.nih.gov/>). However, the genomes of “model” organisms have been finished (or will soon be finished), including the nematode worm (39), the fruit fly (40), the mouse (unpublished but in the public database), other mammals (e.g. dog, rat, chimp, bovine), and plants (e.g. rice, canola). The availability of these genome sequences will support basic biologic studies in each respective organism, as well as provide a better understanding of mechanisms of human disease when they are being modeled in these systems.

The immediate medical application from this wealth of new genomic information is the identification of genes and proteins associated with disease, the pursuit of new diagnostics, and treatments for these diseases based on the elucidated biochemical pathways (Fig. 2). Approximately 5000 human diseases are known to have a genetic component. One thousand disease-associated markers or genes have already been isolated (41), 940 of which exist as available gene-based tests (see <http://www.geneclinics.org/>). Until the 1980s, the primary strategy for identifying human genetic disease genes was to focus on biochemical and physiologic differences between normal and affected individuals. However, for the vast majority of inherited single-gene disorders in human, adequate biochemical

information was either insufficient or too complicated to give insight into the basic biologic defect. In the early 1980s, an alternative paradigm was introduced that suggested disease genes could be isolated solely on their location along the chromosome (42). This approach, now called positional cloning, consists of first determining which of the 23 chromosomes the disease gene resides on (43). Genes on the chromosome are then systematically tested for DNA sequence changes (or mutations) that occur only in the family members having the disease and not found in unaffected individuals. Before the initiation of the HGP, the disease genes for Duchenne/Becker muscular dystrophy and cystic fibrosis, and a few others, were identified. The development of new mapping resources and technologies and now DNA sequences generated by the genome project greatly simplified the positional cloning process of genes and the identification of proteins and their pathways. This has accelerated the pace of discovery of new disease genes including those for fragile X syndrome, Huntington disease, inherited breast cancer, inflammatory bowel disease, early-onset Alzheimer disease, and many others. The rapid advance in genomics has also set the stage for the next paradigm shift in biomedical research: systems biology and multidisciplinary medical research (Fig. 3) (44–46).

With the HGP sequence now well advanced, it becomes possible to dissect the molecular genetics of multifactorial diseases such as cancer and neurodevelopmental and cardiovascular disease. All can involve multiple combinations of genes and their interaction with possibly strong environmental components (47–50). The proper study of these common diseases will require careful large-scale patient and family data collection, thorough clinical examination, systematic DNA analysis, and complex statistical modeling (51). This was nicely demonstrated in the recent mapping of an asthma gene to chromosome 7 (52). Already there have been such significant advances in the development of genotyping (53) and microarray (54, 55) technologies that within 5–10 y it might not be so unreasonable at birth or autopsy to resequence an individual’s genome to extract medically relevant information (29, 56, 57). With all of the rapid progress in the field, any outright optimism is tempered by the fact that only a handful of multifactorial disease genes have been accurately mapped and, even less successfully, identified. The most notable examples

Figure 2. Applications of genome project information. The immediate application of human genome information is the identification of genes and proteins (as well as the pathways they interact in) associated with disease. The earliest application after disease gene discovery is the development of diagnostic tests followed by potential for carrier screening. From information about the biology of the disease, the pursuit of new treatments can be pursued. [Modified from Scherer and Tsui (60).]

Figure 3. Paradigm shifts in biomedical research caused by advances in genomics. Progress in genomics has set the stage for a shift from a “bottom-up” or hypothesis-driven scientific approach to large-scale and multidisciplinary strategies to study more-complex phenomena and systems. [Modified with permission from Peltonen and McKusick (44).]

of success are the recent identification of the calpain-10 (58) and *NOD2* (59) genes associated with type 2 diabetes and Crohn's disease, respectively.

There have also been some far-reaching predictions that DNA-based presymptomatic susceptibility testing for complex genetic diseases will soon be available for practitioners to screen and identify patients at risk (similar to what is available for many monogenic diseases; Fig. 2) (45, 46). Although this will be the case for a few highly penetrant genes, it is our position that for most multifactorial diseases the genetic effect of a single susceptibility locus will likely be small, may be common, and will have low sensitivity and specificity in population-based screening. The greatest utility of discovering these genes will likely be in the identification of biochemical pathways underlying pathology, which in turn will provide a target for therapeutic intervention (Fig. 2). Moreover, as each component of a pathway becomes connected to another, the rules governing the molecular and cellular control of embryogenesis will be better mapped out. Ultimately the genome "parts-list" will extend beyond the gene and protein to include the regulatory circuitry that dictates the hierarchy of mRNAs and proteins that control coordinated embryonic and adult development.

So successful has the human and other genome projects been at generating vast amounts of biologic data that many traditional fields have followed suit by adopting the "-omics" suffix. Genetics and *genomics* are already being used interchangeably as the shift of focus in the former moves from the study of monogenic diseases to complex diseases (Fig. 3). Large-scale cataloging of RNA expression is *transcriptomics*, deciphering proteins and pathways is *proteomics*, the interaction of genes, drugs, and environment is *pharmacogenomics*, physiology and metabolism when necessary have become *physiologomics* and *metabolomics*, and so on. This just might be the true legacy of the HGP: a transition in biomedical research from studying single molecules to the understanding of the natural history of all of the entities in a genome, pathway, organelle, cell, tissue, or organism (Fig. 3). Coupled with computationally intense modeling, more-complete hypotheses will be developed before finally returning to experimental testing in the wet-bench laboratory, model organism, or patient.

REFERENCES

- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huseon DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkuch C, Pratt E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hattori T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X 2001 The sequence of the human genome. *Science* 291:1304-1351
- International Human Genome Sequencing Consortium 2001 Initial sequencing and analysis of the human genome. *Nature* 409:860-921
- Hogenesch JB, Ching KA, Batalov S, Su AI, Walker JR, Zhou Y, Kay SA, Schultz PG, Cooke MP 2001 A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell* 106:413-415
- Modrek B, Lee C 2002 A genomic view of alternative splicing. *Nat Genet* 1:13-19
- Emanuel BS, Shaikh TH 2001 Segmental duplications: an "expanding" role in genomic instability and disease. *Nat Rev Genet* 2:791-800
- Richards RI 2001 Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Hum Mol Genet* 20:2187-2194
- Kazazian HH 1998 Mobile elements and disease. *Curr Opin Genet Dev* 3:343-350
- St George-Hyslop PH 1999 Molecular genetics of Alzheimer disease. *Semin Neurol* 4:371-383
- Samonte RV, Eichler EE 2002 Segmental duplications and the evolution of the primate genome. *Nat Rev Genet* 1:65-72
- Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE 2001 Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res* 6:1005-1017
- Scherer SW, Cheung J 2001 Discovery of the human genome sequence in the public and private databases. *Curr Biol* 20:R808-R811
- Kurotaki N, Imazumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita HA, Kinoshita A, Mizuguchi T, Yoshiura KI, Ohta T, Kishino T, Fukushima Y, Niikawa N, Matsumoto N 2002 Haploinsufficiency of *NSD1* causes Sotos syndrome. *Nat Genet* 4:365-366
- Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC 2001 The AZ Fc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* 3:279-286
- Gratacos M, Nadal M, Martin-Santos R, Pujana MA, Gago J, Peral B, Armengol L, Ponsa I, Miro R, Bulbena A, Estivill X 2001 A polymorphic genomic duplication on human chromosome 15 is a susceptibility factor for panic and phobic disorders. *Cell* 3:367-379
- Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, Costa T, Grebe T, Cox S, Tsui LC, Scherer SW 2001 A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. *Nat Genet* 3:321-325
- Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O 2001 Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. *Am J Hum Genet* 4:874-883
- Giglio S, Calvari V, Gregato G, Gimelli G, Camanini S, Giorda R, Ragusa A, Guerneri S, Selicorni A, Stumm M, Tonnes H, Ventura M, Zollino M, Neri G, Barber J, Wiczorek D, Rocchi M, Zuffardi O 2002 Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation. *Am J Hum Genet* (in press)
- Cai WW, Mao JH, Chow CW, Damani S, Balmain A, Bradley A 2002 Genome-wide detection of chromosomal imbalances in tumors using BAC microarrays. *Nat Biotechnol* 4:393-396
- Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG 1998 High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 2:207-211
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willet DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schmitt B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altschuler D 2001 A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933
- Leeder JS 2001 Pharmacogenetics and pharmacogenomics. *Pediatr Clin North Am* 3:765-781
- Lazarou J, Pomeranz BH, Corey PN 1998 Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 15:1200-1205
- Roses AD 2000 Pharmacogenetics and the practice of medicine. *Nature* 405:857-865
- Evans WE, Relling MV 1999 Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286:487-491
- Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J 1996 A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152-154

26. Kong A, Gudbjartsson DF, Sainz J, Jonsson GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K 2002 A high-resolution recombination map of the human genome. *Nat Genet* 3:241–247
27. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES 2001 High-resolution haplotype structure in the human genome. *Nat Genet* 2:229–232
28. Helmut L 2001 Genome research: map of the human genome 3.0. *Science* 293:583–585
29. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR 2001 Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294:1719–1723
30. O'Brien SJ, Menotti-Raymond M, Murphy WJ, Nash WG, Wienberg J, Stanyon R, Copeland NG, Jenkins NA 1999 The promise of comparative genomics in mammals. *Science* 286:458–62
31. Pennacchio LA, Rubin EM 2001 Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* 2:100–109
32. Owens K, King M-C 1999 Genomic views on human history. *Science* 286:451–455
33. Paabo S 2001 Genomics and society: the human genome and our view of ourselves. *Science* 291:1219–1220
34. Doolittle WF 1999 Lateral genomics. *Trends Cell Biol* 12:5–8
35. Salzberg SL, White O, Peterson J, Eisen JA 2001 Microbial genes in the human genome: lateral transfer or gene loss? *Science* 292:1903–1906
36. Mushegian AR, Koonin EV 1996 A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* 93:10268–10273
37. Broder S, Venter JC 2000 Whole genomes: the foundation of new biology and medicine. *Curr Opin Biotechnol* 11:581–585
38. Subramanian G, Mural R, Hoffman SL, Venter JC, Broder S 2001 Microbial disease in humans: a genomic perspective. *Mol Diagn* 4:243–252
39. The *C. elegans* Sequencing Consortium 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012–2018
40. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC 2000 A whole-genome assembly of *Drosophila*. *Science* 287:2196–2204
41. Antonarakis SE, McKusick VA 2000 OMIM passes the 1,000-disease-gene mark. [letter] *Nat Genet* 25:11
42. Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Cole FS, Curnutte JT, Orkin SK 1986 Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location. *Nature* 322:32–38
43. Collins FS 1992 Positional cloning: let's not call it reverse anymore. *Nat Genet* 1:3–6
44. Peltonen L, McKusick, VA 2001 Dissecting human diseases in the postgenomic era. *Science* 291:1224–1229
45. Collins FS, McKusick VA 2001 Implications of the Human Genome Project for medical science. *JAMA* 285:540–544
46. Subramanian G, Adams MD, Venter JC, Broder S 2001 Implications of the human genome for understanding human biology and medicine. *JAMA* 286:2296–2307
47. Cravchik A, Subramanian G, Broder S, Venter JC 2001 Sequence analysis of the human genome: implications for the understanding of nervous system function and disease. *Arch Neurol* 11:1772–1778
48. Zhang W, Laborde PM, Coombes KR, Berry DA, Hamilton SR 2001 Cancer genomics: promises and complexities. *Clin Cancer Res* 8:2159–2167
49. Rees J 2002 Complex disease and the new clinical sciences. *Science* 296:698–701
50. Strohm R 2002 Maneuvering in the complex path from genotype to phenotype. *Science* 296:701–703
51. Willet WC 2002 Balancing life-style and genomics research for disease prevention. *Science* 296:695–698
52. Laitinen T, Daly MJ, Rioux JD, Kauppi P, Laprise C, Petais T, Green T, Cargill M, Hahtela T, Lander ES, Laitinen LA, Hudson TJ, Kere J 2001 A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat Genet* 1:87–91
53. Brennan MD 2001 High throughput genotyping technologies for pharmacogenomics. *Am J Pharmacogenomics* 4:295–302
54. Golub TR 2001 Genomic approaches to the pathogenesis of hematologic malignancy. *Curr Opin Hematol* 4:252–261
55. Triche TJ, Schofield D, Buckley J 2001 DNA microarrays in pediatric cancer. *Cancer J* 1:2–15
56. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP 1996 Accessing genetic information with high-density DNA arrays. *Science* 274:610–614
57. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR 2002 Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296:916–919
58. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI 2000 Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2:163–175
59. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G 2001 Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411:599–603
60. Scherer SW, Tsui LC 2001 The Human Genome Project. In: Sensen CW, Reh HJ, Reed G (eds) *Biotechnology: Genomics and Bioinformatics*, Vol 5b. Wiley-VCH Publishers, Weinheim, Germany, pp 42–56