

### *Mendelian Inheritance in Man* and Its Online Version, OMIM

Victor A. McKusick

Last year marked the 40th anniversary of the publication of the first print edition of *Mendelian Inheritance in Man* (MIM).<sup>1</sup> This seems an appropriate juncture at which to review its origins, evolution, and present status, including and particularly those of its online version, OMIM (Online Mendelian Inheritance in Man). This is an opportunity, at the same time, to review in brief the rapid progress in an important part of medical genetics and genomics, as chronicled in MIM/OMIM over these 40 years, and to contemplate the future challenges of OMIM.

#### Description of MIM/OMIM

MIM<sup>1,2</sup> is a comprehensive knowledgebase of human genes and genetic disorders. It consists of full-text overviews of genes and genetic phenotypes, particularly disorders, and is useful to students, researchers, and clinicians. It was initiated in the early 1960s as a trilogy of catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes. It has been maintained as an electronic file since 1964 and has been published in 12 print editions (fig. 1), the first in 1966, the most recent (in three volumes) in 1998. (Various editions of MIM were translated into Russian [1976], Spanish [1976], and Mandarin [1996]; see fig. 2.) In 1987, it became generally available on the Internet, under the designation "OMIM," from the Welch Medical Library at Johns Hopkins University. Since December 1995, it has been distributed on the World Wide Web from the National Center for Biotechnology Information (NCBI) of the National Library of Medicine. This knowledgebase is updated daily. Authoring and editing are headquartered at Johns Hopkins University School of Medicine.

#### Origins and Evolution of the Structure and Organization and Content of MIM/OMIM

MIM had its origins in three different endeavors<sup>3</sup>: first, the annual reviews of medical genetics that my colleagues and I in the Moore Clinic at Johns Hopkins prepared for each of 6 years, 1958–1963<sup>4,5</sup>; second, a catalog of X-linked traits that I compiled in 1962 as an assessment of the genetic content of the X chromosome<sup>6,7</sup>; and, third, a catalog of autosomal recessive phenotypes that I prepared in

1963 as a resource for identifying both "old" and "new" recessive diseases in studies of the Old Order Amish.<sup>8,9</sup>

In the original three catalogs corresponding to the three major modes of Mendelian inheritance, the entries were arranged alphabetically according to the preferred title of the particular phenotype, and numbering was done consecutively. By the 11th book edition in 1994,<sup>2</sup> the X-linked catalog had been joined by two other chromosome-specific catalogs: those for the Y chromosome and for the mitochondrial chromosome. Entries in the original autosomal dominant, autosomal recessive, and X-linked catalogs had been assigned unique identification numbers, beginning with 1, 2, and 3, respectively. Entries in the two new chromosome-specific catalogs were given unique numbers beginning with 4 (Y-linked) and 5 (mitochondrial). (Entries were given 4-digit numbers in the first [1966] and second [1968] book editions of MIM. Entry numbers were expanded to 5 digits with the third [1975] edition by adding a zero to the 4-digit numbers of previously existing entries and to 6 digits in the 9th [1990] edition by the same method.)

Since May 1994, a distinction between autosomal dominant and autosomal recessive traits has not been maintained in MIM. All autosomal traits (or genes) for which new entries were created were added consecutively to a new catalog, with 6-digit numbers beginning with 6. No new entries were added to the original autosomal dominant and autosomal recessive catalogs that had unique entry numbers beginning with 1 and 2, although copious new information bearing on previously existing entries in these original catalogs has been added. (A caveat: some entries that remain in the catalogs with numbers beginning with 1 or 2 relate to phenotypes not now considered dominant or recessive, respectively, in the light of newer understanding.)

The reasons for discontinuing the distinction between autosomal dominant and autosomal recessive entries included the fact that entries were being created for an increasing number of genes for which there was extensive information, including location on a specific autosome, but no associated Mendelian phenotypic variation with either dominant or recessive inheritance. Also, the distinction is only relative—that is, whether dominant or recessive sometimes depends on the level at which the phe-

From the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore  
Received January 31, 2007; accepted for publication February 15, 2007; electronically published March 8, 2007.

Address for correspondence and reprints: Dr. Victor A. McKusick, 1007 Blalock Building, Johns Hopkins Hospital, 600 N. Wolfe Street, Baltimore, MD 21287-4922. E-mail: mckusick@peas.welch.jhu.edu  
*Am. J. Hum. Genet.* 2007;80:588–604. © 2007 by The American Society of Human Genetics. All rights reserved. 0002-9297/2007/8004-0003\$15.00  
DOI: 10.1086/514346



**Figure 1.** Twelve print editions of MIM, the first published in 1966 and the most recent, in three volumes, published in 1998

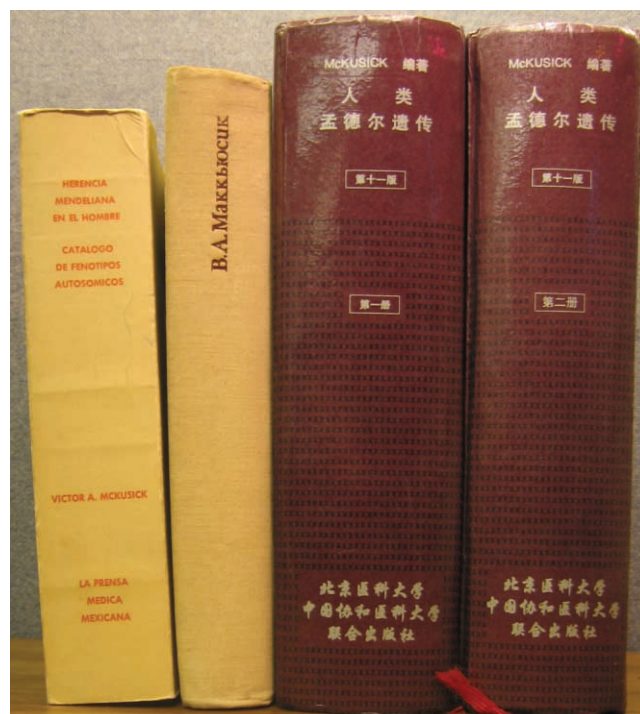
notype is analyzed. For example, in several of the red-cell enzymopathies, the deficiency state is autosomal recessive but the electrophoretic variation is likely to be demonstrable in the heterozygote—that is, it is dominant, or at least intermediate, in its inheritance. Furthermore, there are rather numerous examples of particular phenotypes that are inherited as dominant or recessive based on different mutations in the same gene. Table 15 on page 1 in the preface of MIM12<sup>2</sup> listed 11 disorders that have both dominant and recessive forms resulting from different mutations in the same gene. Some mutations that cause absent function of the protein—that is, null mutations—produce phenotypic effects only with homozygosity; missense mutations may cause the disorder because of a dominant negative effect when, for example, the structure of a protein that is part of a heteromeric protein complex is altered.

Starting with the first edition of MIM, two classes of entries were differentiated. Those for which the particular mode of inheritance was considered quite certain and the phenotype was thought to be distinct from any already represented by an entry were distinguished by an asterisk (\*) preceding the unique entry number. The asterisk was omitted in the case of entries for which the phenotype was less certainly Mendelian, less clearly of the particular mode of inheritance, or not distinct from a phenotype described in another entry. The inclusion of these unasterisked entries was considered important for heuristic purposes.

From the beginning, the gene behind the phenotype was always kept in mind. For most of the first 15 years of the catalogs and more, however, there was little way to know whether a given phenotype was in fact caused by mutation in a single gene or might be caused by mutation in any one of two or more different genes, and it was usually impossible to tell whether two or more quite different phenotypes were due to mutations in the same gene. Since 1990, separate entries have been, as a rule,

created for phenotypes and the genes that have mutations causing those phenotypes.

The practice of separate entries for genes and phenotypes was initiated, in large part, to handle the issues of one phenotype—several genes and of one gene—several phenotypes. Entries describing phenotypes for which the mutational basis has been found in one or more genes are flagged with a number sign (#) preceding the unique entry



**Figure 2.** Foreign-language editions of MIM: (left to right) Spanish (Mexican) edition, translated by Rudolfo Guzmán Toledano, 1976; Russian edition, translated by E. K. Gentera and V. I. Ivanova, 1976; and Mandarin edition, translated by Wilson H. Y. Lo and others (two volumes), 1996.

number, and the initial paragraph indicates the entry number of the gene(s) in which the mutation(s) is described. Gene entries have an asterisk preceding the unique number. Entries that contain both phenotype and gene information are flagged with a plus sign (+); X-linked examples include HPRT (MIM +308000), G6PD (MIM +305900), and HEMA (MIM +306700). Beginning in 2004, other Mendelizing phenotypes, regardless of whether they have been mapped, have been denoted by a percent sign (%) preceding the entry number when the causative gene has not yet been identified and cloned. Table 1 presents current statistics on these several categories of entries.

Beginning with hemoglobinopathies as early as the first edition (1966) and in full force by 1988 (MIM8), allelic variants (AVs) (or mutations) have been appended to the gene entries—for example, the beta-globin gene (*HBB*) entry (MIM +141900). At present, each AV is given a unique 10-digit number, consisting of the primary 6-digit number for the gene followed by a 4-digit extension beginning with .0001 for the first listed AV. As of December 4, 2006, the *HBB* entry cataloged 537 AVs of the *HBB* gene, numbered MIM +141900.0001 to +141900.0537. The entry of each AV consists of the title of the trait (phenotype) determined by the mutation, the gene symbol and the shorthand description of the mutation,<sup>10–12</sup> text providing a varying amount of information on the family(ies) or population(s) studied, the details of the specific DNA change, and peculiarities of phenotype and genetics. “Allelic Variants” was selected as the heading of that section of gene entries rather than “Mutations” because, together, they represent an allelic series. Furthermore, the title of each AV is the phenotype (not the mutation), which, in some instances, can be an electrophoretic or antigenic polymorphism of an enzyme or plasma protein. The molecular bases of the variation in blood-group antigens are given as AVs, for example.

Selection of particular mutations of a given gene for inclusion as AVs in OMIM has been based on general criteria, including the following: (1) the first or first few disease-related mutations to be identified in the given gene; (2) any mutation with a particularly high frequency, such as Phe508del in the *CFTR* gene (MIM \*602421.0001) in cystic fibrosis (MIM #219700); (3) a mutation related to a distinct phenotype not previously represented in the list; (4) mutations of historical interest, such as the specific mutation in the family or population in which the phe-

notype was first described—for example, the mutation in the *CLCN1* gene (MIM \*118425.0006) in the family of Dr. Thomsen, who first described Thomsen disease (MIM #160800); (5) any mutation with a peculiar ethnic or geographic distribution; (6) any mutation arising through a distinctive mutagenic mechanism such as gene conversion (as in classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency [MIM \*201910.0001]) or gene fusion (as in Hb Lepore [MIM +142000.0019ff]); (7) any mutation producing the phenotype through a distinctive pathogenetic mechanism; (8) mutations associated with autosomal dominant versus autosomal recessive inheritance with different mutations in the same gene, as in the recessive (MIM \*139250.0005) and dominant (MIM \*139250.0007) forms of isolated growth hormone deficiency due to allelic mutations in the growth hormone gene (*GHI*); and (9) polymorphisms demonstrating association with disease—for example, the Y402H polymorphism of complement factor H (*CFH* [MIM \*134370.0008]) in age-related macular degeneration (MIM #603075). The last category represents so-called susceptibility genes, as discussed later. Most are part of the multifactorial basis of common disorders.

OMIM provides links to comprehensive mutation listings, including the Human Gene Mutation Database curated at Cardiff, and many locus-specific mutation databases (LSDBs)—for example, PAHdb (Phenylalanine Hydroxylase Locus Knowledgebase) and CFTRdb (Cystic Fibrosis Mutation Database) for PKU and cystic fibrosis mutations, respectively, as well as the Human Genome Variation Society based in Melbourne, which maintains information on >500 LSDBs.

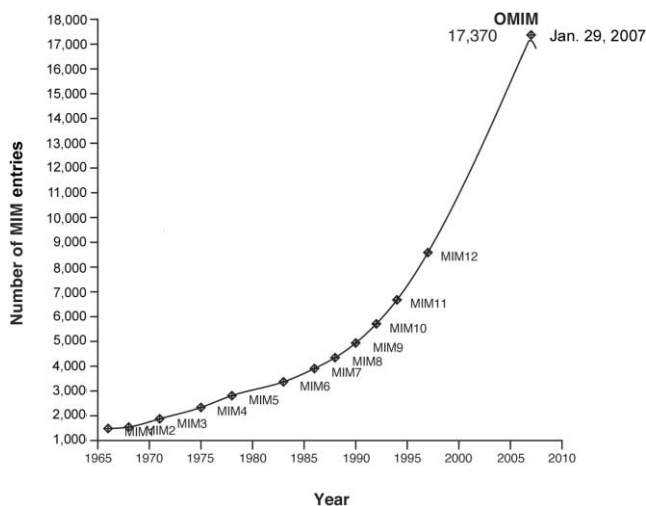
### Evolution of MIM/OMIM from a Catalog of Mendelian Phenotypes to a Catalog of Human Genes and Genetic Disorders

The first print edition of MIM in 1966 had the subtitle *Catalogs of Autosomal Dominant, Autosomal Recessive and X-linked Phenotypes*.<sup>1</sup> In the 1994 edition,<sup>2</sup> the subtitle became *A Catalog of Human Genes and Genetic Disorders*—a reflection of the progress in the field since the 1960s.

Nearly all of the 1,486 entries in the first edition of MIM discussed phenotypes. As of January 29, 2007, 6,146 of the >17,300 entries in OMIM represented phenotypes (see table 1 and fig. 3); the rest related to genes. Beginning in

**Table 1. OMIM Statistics as of January 29, 2007**

Entry Classification	No. of Entries by Category				
	Autosomal	X Linked	Y Linked	Mitochondrial	Total
* Gene with known sequence	10,644	495	48	37	11,224
+ Gene with known sequence and phenotype	356	32	0	0	388
# Phenotype description, molecular basis known	1,851	169	2	26	2,048
% Mendelian phenotype or locus, molecular basis unknown	1,411	134	4	0	1,550
Other, mainly phenotypes with suspected Mendelian basis	<u>2,014</u>	<u>144</u>	<u>2</u>	<u>0</u>	<u>2,160</u>
Total	16,276	974	56	63	17,370



**Figure 3.** Growth of MIM and OMIM, in terms of total number of entries.

the late 1960s, entries were created in MIM for individual genes for which no associated Mendelian phenotype was known. The method of interspecific (e.g., mouse-human) somatic cell hybridization made it possible to map genes to specific human chromosomes without the existence of a Mendelizing phenotype that could be used in family linkage studies. The difference between the genomes of the two species in the hybrid substituted for the differences between the genomes inherited from father and mother used in family linkage mapping. Thus, when the thymidine kinase gene (*TK1* [MIM \*188300]) was mapped to chromosome 17 by study of mouse-human hybrid cell lines,<sup>13,14</sup> an entry was created for the gene, even though no Mendelian variation was known. Already, separate entries had been created for hemoglobin genes such as HBB, the lactate dehydrogenase genes such as *LDHA* (MIM +150000), and the *G6PD* gene (MIM +305900), among others. As mapping and cloning of genes advanced, the genes involved were given new entries in MIM, again although no Mendelian variation may have been known. Until the generic autosomal catalog was established (in 1994), autosomal “gene entries” were arbitrarily incorporated in the autosomal dominant catalog.

In the accessioning and curating of gene entries, OMIM works closely with the NCBI reference sequence project and the Human Genome Organisation (HUGO) nomenclature committee. Because each group is involved in curating genes and sequence information, a method for sharing each group’s analyses was established under the direction of Donna Maglott at NCBI. This collaborative effort resulted in the public resource “Locus Link” and its successor “Entrez Gene.” This initiative allows Alan Scott, OMIM’s Deputy Scientific Director for Genes, to review genes in OMIM, to remove duplicates, and to identify “new” genes for consideration of inclusion in OMIM. This

is a considerable undertaking, since nearly 23,000 genes with supporting sequence have been identified.

### Evolution of MIM/OMIM as an Electronic Resource

MIM has been maintained on computer since early 1964.<sup>15</sup> In those pre-word processor days, maintenance on the mainframe computer was a boon to the updating process and facilitated preparation of camera copy, including author and subject indices, for book publication. That advantage for book publication continued. For example, from 1986 (MIM7) to 1994 (MIM11), it was possible to produce a print edition in 4 months at 2-year intervals. With closing of the files on March 1, bound books would be available by July 1. A time-consuming task during those 4 months was preparation of the front material (including the Synopsis of the Human Gene Map) and the indices. The process of vetting the indices in preparation for the book was valuable for detection of errors, including duplications, misspellings, inconsistent nomenclature, etc. The book was published by photo-offset of the computer printout (in all uppercase letters for the first three editions, 1966, 1968, and 1971) and of camera-ready copy prepared by automatic typesetting in the subsequent editions.

In the 1980s, a new era began, with the adoption of MIM by the National Library of Medicine as the test bed for the development of IRx (Information Retrieval Experiment), a method of authoring and editing that permitted the rapid search of specific text material.<sup>16</sup> An online version of the 6th edition of MIM (1983) with updates and with the IRx search engine was demonstrated at the Bar Harbor Medical Genetics Course in July 1985 and was used as a resource at the eighth Human Gene Mapping Workshop in Helsinki in August 1985. By the fall of 1985, an online version of MIM, now called “OMIM,” with the IRx search engine became a major aid in authoring and editing OMIM. Searchability helped to avoid duplications and inconsistencies and allowed entries to be related to each other—cross-referenced—more easily. Beginning in September 1987, OMIM was made generally accessible on the Internet from the Welch Medical Library of Johns Hopkins University. The informatics aspects of OMIM were transferred to the NCBI of the National Library of Medicine on December 1, 1995. Thus, in the 31 years from 1964 to 1995, MIM went from a solitary resource on magnetic tape to a cornerstone genetics resource on the World Wide Web, integrated with other primary genetic data sources. OMIM was one of the first electronic resources to exploit the advantages of the Web.

### Nosology and Nomenclature in MIM/OMIM

Nosology (literally meaning “the study of disease” but customarily taken to mean the classification or delineation of disease), nosography (“the description of disease”), and the nomenclature of disease have necessarily been central

considerations throughout the history of MIM. Indeed, nosology and nosography, viewed as the delineation of distinct genetic traits, are the main bases for assembling MIM. The basic premise has been that “Mendelization” indicates that the phenotype represents a distinct disorder or trait related to a specific mutation in a specific gene and deserves a specific name. In his monograph entitled *Nosography*,<sup>17</sup> which was published in 1930, Knut Faber, Copenhagen professor of medicine, attributed to Mendel a major, albeit indirect, role in guiding thinking along lines of specific disease entities with specific etiology (causation), comparable to the role played by the pioneers in bacteriology in the late 1800s and early 1900s. Two hundred years ago, angina pectoris, asthma, consumption, dropsy, jaundice, stroke, and many symptom complexes were viewed and managed as though they were distinct entities with a single etiology.

Nosology and the related naming of genetic disorders have both practical and theoretical significance—practical because of their importance to diagnosis, prognosis, and management and theoretical because of the implication that the disorder represents a distinct entity with a distinct mutational basis. These were important topics of discussion at the five annual conferences on the clinical delineation of birth defects held at the Johns Hopkins Hospital (1968–1972). (The proceedings of these five conferences on the Clinical Delineation of Birth Defects were published by the March of Dimes in 16 volumes. See, for example, the publication on skeletal dysplasias, which were discussed on the 3rd and 4th days of the first conference in 1968.<sup>18</sup>) When a Mendelizing disorder was described in a single large kindred or in a collection of kindreds, the question arose as to whether the condition was the same as a previously described entity or the same in all kindreds studied. The question brought out debates between “lumpers” and “splitters.”<sup>19,20</sup> The lumpers and splitter controversies have often been resolved only by molecular elucidation at the DNA level. It turned out that both the lumpers and the splitters were in part correct. Molecular elucidation revealed numerous instances of “many from one” (multiple phenotypes from different mutations in the same gene) and “one from many” (the same phenotype from mutations in two or more separate genes).

The main principles of clinical genetics are pleiotropism, genetic heterogeneity, and variation. Pleiotropism—multiple phenotypic effects of a single mutant gene, the basis of syndromes—mandates lumping of disorders that may have been separately described on the basis of features that predominate in one group of patients. Genetic heterogeneity, however recognized, is a basis for splitting. Some aspects of variation will be discussed later, in connection with multifactorial inheritance.

With molecular characterization of an increasing number of genetic disorders through gene mapping followed by positional cloning, genetic heterogeneity independent of phenotypic heterogeneity has been recognized as more frequent than previously realized. When the causative mu-

tations are at different genes/loci, OMIM now considers the disorders to be distinct entities. See, for example, the several forms of long QT syndrome: LQT1 (MIM #192500), LQT2 (MIM +152427), LQT3 (MIM #603830), LQT4 (MIM #600919), LQT5 (MIM +176261), LQT6 (MIM +603796), and LQT7 (MIM #170390). Often, the phenotype can be discerned to be somewhat different when the molecular characterization permits the study of “pure-culture” groups of cases. Additionally, treatment of the disorder may be found to be specific to the underlying genetic basis. “Many from one,” different phenotypes with different mutations in the same gene, is illustrated by many examples. The multiple disorders resulting from mutations in *HBB*, a total of at least seven, represents an early example. The multiple disorders resulting from mutations in the lamin A/C gene (*LMNA* [MIM \*150330])—a total of at least 11, varying from progeria (MIM #176670) to a form of Charcot-Marie-Tooth disease (MIM #605588)—provide an impressive recent example.

In general, usage has dictated the choice of the preferred designations for disorders. The preferred designations change over time, as popular usage changes; “mucoviscidosis” became “cystic fibrosis of the pancreas” and then simply “cystic fibrosis.” Usage (in the choice of particular eponyms, for example) tended to vary some between the United States and Europe, but this is now much less the case. The preferred designations in OMIM have increasingly been used as a standard in publications, along with the inclusion of OMIM entry numbers. All alternative designations of both phenotypes and genes, including historical ones, are usually listed in the heading of each entry.

The initial choice of names for “new” genetic disorders has been chaotic. In some instances, one feature of a pleiotropic syndrome was selected to designate the whole; examples include arachnodactyly for Marfan syndrome (MIM #154700) and angiokeratoma for Fabry disease (MIM #301500). However, often, over time, a constellation of features was found to be more important to clinical diagnosis than was a single feature. Partly for this reason, eponymic designations have been used heavily and have much to recommend them. They have the advantage of conveying no preconceived notion as to the basic nature of the abnormality. “Hurler syndrome” was, time showed, a better designation than was “lipochondrodystrophy,” which Index Medicus continued to use long after the fundamental fault was found to concern mucopolysaccharides, not lipid. The eponym is merely a “handle.” Often the person whose name is used was not the first to describe the disorder or did not describe the full syndrome as it subsequently came to be known. Priority disputes I have never considered important; when it is found that Jones in fact described a disorder before Smith, it seems pointless to change from a widely established use of the Smith eponym.

MIM/OMIM has consistently used the nonpossessive form of eponyms—for example, “Marfan syndrome,” not “Marfan’s syndrome.” In this practice, I was schooled by

J. Earle Moore and others who edited the first edition (1956) of my *Heritable Disorders of Connective Tissue*<sup>21</sup> and used the then-current edition of the *AMA Manual of Style*. Among other advantages, use of the nonpossessive avoids the mistake of putting an apostrophe in the wrong place, writing, for example, “Wilm’s tumor” for “Wilms tumor” and “Grave’s disease” for “Graves disease.”

It has been useful in speech and writing to have acronyms pronounced as single words, for example, TAR syndrome (MIM #274000) for thrombocytopenia–absent radius syndrome, and VATER association (MIM 192350) for vertebral defects, anal atresia, tracheoesophageal fistula with esophageal atresia, and radial dysplasia. Initialisms, such as OFD1 (MIM #311200)–OFD10 (MIM 165590) for orofacidigital syndromes, have also been useful. Family initials, however, as in the Opitz G/BBB syndrome (MIM #300000), have been abandoned as a method of naming.

Geographic designations are sometimes used—for example, familial Mediterranean fever (MIM #249100 and MIM #134610) and Tangier disease (MIM #205400), but these can prove a problem when populations outside the originally described regions are found to have the disease. In some cases, the nature of the basic defect is used as the name. Factor VIII deficiency (hemophilia A [MIM +306700]) and G6PD deficiency (MIM #305900) are examples. In some ways, this has advantages, but the disadvantage is indicated by the absence of clinical specificity; there are, for example, a considerable number of different forms of G6PD deficiency, in terms of clinical presentation.

Pejorative or demeaning designations have always been discouraged.<sup>22–24</sup> They should never be used in the clinic. “Gargoyle” (or “gargoylism”) for “Hurler syndrome” (MIM #607014) is a case in point; “bulldog syndrome” (MIM #312870) is another. Feingold<sup>22</sup> appropriately objected to “bird-headed dwarf” (MIM #210600), “Michelin tire baby syndrome” (MIM #156610), and “happy puppet syndrome” (MIM #105830), none of which is a preferred designation in OMIM. Cohen<sup>23</sup> pointed out a possible problem with gene names in communicating to laymen the basic nature of the form of holoprosencephaly (HPE3 [MIM #142945]) due to mutations in the “sonic hedgehog” gene (SHH [MIM #600725]). Ludman<sup>24</sup> stated that he dreads the day when he will need to counsel a family with a child with spondylocostal dysostosis (MIM #609813) caused by mutations in the “lunatic fringe” gene (*LFNG* [MIM #602576]). Actually, I doubt that these bizarre names, which are derived from their homologs in *Drosophila*, will represent a serious problem. The gene symbol can be used if that seems preferable.

As referred to earlier, a numbering system has also come into extensive use in connection with disorders that are identical or nearly identical but that are due to mutations in separate genes. In the past decade or so, OMIM has made increasing use of numerical or alphabetic systems for phenotypic series (e.g., DFNB#- for the many [ $>68$ ] forms of autosomal recessive nonsyndromic deafness,

LQT#- for the many [ $>7$ ] forms of long QT syndrome, and SPG#- for the many [ $>33$ ] forms of spastic paraplegia). Historically, the precedent for numbering phenotypic whi series was established with the glycogen-storage diseases, the mucopolysaccharidoses, and the Ehlers-Danlos syndromes. Confusion can arise when workers use different numbers for the same disorder (e.g., see the complementation groups of peroxisomal disorders [MIM +170993ff]) or if the phenotype was not fully characterized before being entered into the series.

Nomenclature is important, not only in the designation of genetic disorders, but also in the naming of genes and the derivation of appropriate gene symbols.<sup>25</sup> Designations for genes and gene symbols based on those names have been the responsibility, for  $>30$  years, of the Gene Nomenclature Committee of the Human Gene Mapping Workshops held every year or two, beginning in 1973, and of its successor, the Human Genome Nomenclature Committee (HGNC) of HUGO, which assumed the responsibility in 1991.

The naming of the gene and gene product in the case of “novel” genes found to be mutant in “mystery” diseases, often through mapping followed by positional cloning, has been based in many instances on the name of the clinical disorder: dystrophin (DMD [MIM #300377]) for the gene product defective in Duchenne muscular dystrophy (MIM #310200) was an early example. Huntingtin (MIM \*143100), neurofibromin-1 (MIM \*162200) and -2 (MIM \*101000), and emerin (MIM \*300384) are others.

### Connecting Phenotypes with Genes in MIM/OMIM

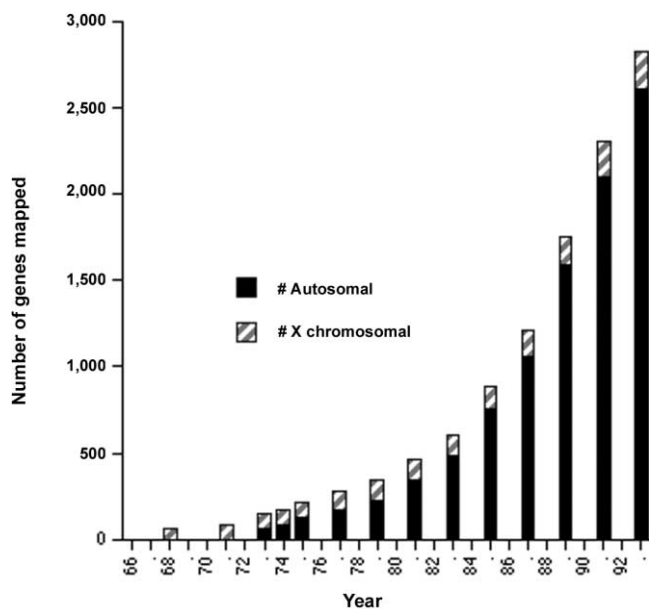
Genetics was defined by William Bateson as the science of biologic variation. It can also be defined simply as the study of inheritance, and genomics as the study of genomes.<sup>26</sup> However genetics is defined, one of its main objectives is to identify specific genetic elements underlying specific phenotypes—to connect phenotype and gene. The first of the three original phenotype catalogs, that for the X chromosome, was published in 1962<sup>6</sup> as one of three parts of a review “On the X chromosome of Man.” The catalog of X-linked traits was assembled as an assessment of the gene content of the X chromosome. The list of traits, most of them disorders or diseases, was compared to a photographic negative from which a positive picture of the genetic constitution of the X chromosome could be derived. Although the second catalog, that for autosomal recessive traits, was assembled for utilitarian purposes—that is, as a resource in the identification of new recessive diseases in inbred groups such as the Amish—the “gene behind the phenotype” was always in mind, and no more than one entry per gene was wittingly made from the beginning of MIM in the 1960s. However, in the age before human molecular genetics, the one gene–several phenotypes and one phenotype–several genes complexity meant that the one gene–one entry rule was often on shaky ground.

Connecting phenotype to gene is particularly important for

medical genetics, because definition of the precise mutational lesion allows specific diagnosis. Moreover, it separates homogeneous clusters of cases for evaluation of prognosis and therapy and for elucidation of the steps from gene back to phene that can be important to therapy and prevention.

Mapping the chromosomes—that is, defining the locus of the genetic elements responsible for particular phenotypes to specific chromosomal sites—was a first step in connecting phene to gene. The mapping process began for the autosomes in 1968<sup>27</sup> and advanced rapidly in the next 20 years (fig. 4). (In 1968, as indicated by the second edition of MIM published that year, 68 phenotypes, judged to be clearly X linked and therefore honored with an asterisk, had been mapped to the X chromosome, but the regional localization on the X chromosome had not yet been established for any of them.)

As of January 29, 2007, OMIM contained information about 2,002 genes that had at least one disease-related mutation. (That count had reached 1,000 by January 1, 2000.<sup>28</sup>) These 2,002 are the gene entries with at least one AV. Because many genes have more than one distinct phenotype in their mutational repertoire, the total number of phenotypes that have been tracked to the DNA level was 3,345 (table 2). (As noted earlier, this count treats the phenotype as separate if it is caused by a mutation in a different gene.) Thus, on average, 1.7 phenotypes have



**Figure 4.** Growth of information in MIM concerning mapping of genes and genetic loci to specific human chromosomes in the period up to the initiation of the Human Genome (Sequencing) Project. In 1968, when the first gene was assigned to a specific autosome (the Duffy blood group to the centromeric region of chromosome 1),<sup>27</sup> MIM recorded 68 X-linked phenotypes with an asterisk, indicating confidence in X-linked inheritance.

**Table 2. Mapping of Clinical Disorders (January 29, 2007)**

Mapping Type	No. Mapped
Loci associated with disorders	3,003
Disorders mapped by association with the gene product	159
Disorders mapped by linkage	942
Disorders "molecularized"	3,345 <sup>a</sup>
Total no. of disorders mapped	4,446

<sup>a</sup> Number of phenotypes labeled with "(3)" in the Disorder field of the Synopsis of the Human Gene Map.

been related to each of the 2,002 genes. Most of this connecting of phene with specific gene has occurred in the past 20 years. Progress during those years was graphed by Peltonen and McKusick.<sup>29</sup> Much of it has been achieved by gene mapping (chromosomal mapping of the locus for the phenotype, to be specific), followed by positional cloning of a previously unknown gene or by the positional candidate-gene approach.

It came as something of a surprise when the complete sequence of the human genome revealed many fewer genes, perhaps by a factor of 10, than might be predicted from the abundance of gene products. That a gene may be subject to mutations that cause diverse phenotypes has a cognate phenomenon: a gene may encode a diversity of gene products through mechanisms of combinatorial alternative splicing, posttranslational modification, and other mechanisms. OMIM attempts to include information about the mechanism of the divergent pathologic phenotypes that occur from mutations in a single gene.

The near ultimate in genotype/phenotype correlation is provided by a few examples in which change in a single codon of a single gene results in the disorder. The phenotype in each case is as stereotypic as the genotype; mutations in other regions of the same gene result in different phenotypes. Striking examples include achondroplasia (MIM #100800) due to the c.1138G→A (Gly380Arg) mutation in the *FGFR3* gene (MIM \*134934.0001), Hutchinson-Gilford progeria syndrome (MIM #176670) due to the c.1824C→T (Gly608Gly) mutation in the *LMNA* gene (MIM \*150330.0022), and fibrodysplasia ossificans progressiva (MIM #135100) due to the c.617G→A (Arg206His) mutation in the *ACVRI* gene (MIM \*102576.0001). In the case of achondroplasia, a few cases are due to a different nucleotide substitution in the same codon: c.1138G→C (Gly380Arg [MIM \*134934.0002]). Similarly, a mutation in the same codon as is involved in the majority of cases of progeria has been found in some cases of that disorder: G→A (Gly608Ser [MIM \*150330.0023]).

In ~70% of cases, Apert syndrome (MIM #101200) is caused by a c.934C→G (Ser252Trp) mutation in the *FGFR2* gene (MIM \*176943.0010) and, in rare other cases, by a mutation in the next adjacent codon, c.937C→G (Pro253-Arg [MIM \*176943.0011]). No phenotypic features distinguishing the two genotypic forms were found.<sup>30</sup> As cataloged in OMIM, the *FGFR3*, *LMNA*, and *FGFR2* genes are

also the sites of mutations causing, respectively, 9, 12, and 11 other phenotypically distinct disorders.

## A Garrodian Perspective on MIM/OMIM

*"The Lessons of Rare Maladies"*

In 1956, I dedicated the first edition of my *Heritable Disorders of Connective Tissue*<sup>21</sup> to Archibald Garrod "and to all who believe, as he did, that the clinical investigation of hereditary disorders can shed light on normal developmental and biochemical mechanisms." (The dedication was accompanied by a previously unpublished etching of Garrod in academic garb by T. Binney Gibbs [created in 1922]. This was provided to me by Garrod's daughter, distinguished Cambridge University archeologist Dorothy A. E. Garrod.) The preface of *Heritable Disorders of Connective Tissue* reproduced the now-well-known quotation from a letter written (in Latin) by William Harvey in 1657 that Garrod included (in translation) in his paper entitled "The lessons of rare maladies" published in *Lancet* in 1928<sup>31</sup>:

Nature is nowhere accustomed more openly to display her secret mysteries than in cases where she shows traces of her workings apart from the beaten path; nor is there any better way to advance the proper practice of medicine than to give our minds to the discovery of the usual law of nature by careful investigation of cases of rarer forms of disease. For it has been found, in almost all things, that what they contain of useful or applicable nature is hardly perceived unless we are deprived of them, or they become deranged in some way.

This Harveian/Garrodian principle has been extensively documented in the case of rare genetic syndromes. The many rare disorders cataloged in OMIM are "experiments of nature" with much to teach about normal biochemical, developmental, and physiologic mechanisms, and indeed much has been learned from them, especially in the 20+ years since the first "disease gene" identified by positional cloning was recorded in OMIM.

Increasingly, basic scientists turn to the human for exploration of the significance of findings in experimental systems or look for "human models" of phenotypes or phenomena in *Caenorhabditis elegans*, *Drosophila*, mouse, and other experimental species. The researcher asks, "Has a defect related to 'my' gene or protein been identified in the human?" OMIM has proved a useful way to find human models of "disorders" in experimental organisms.

A human-interest story in this connection involves the late Robert J. Gorlin (1923–2006) and his son Jed B. Gorlin. Jed cloned the filamin A gene (*FLNA* [MIM +300017]) in 1990<sup>32</sup> and mapped it to Xq28 in 1993.<sup>33</sup> It was of particular delight to his father when *FLNA* was found (by others) to be the site of mutations underlying frontometaphyseal dysplasia (FMD [MIM #305620]), otopalatodigital syndrome (OPD1 [MIM #311300]), and several other disorders for which the father had provided definitive clinical descriptions as well as names.

## Garrod's Generalization: Most Diseases Are Related to Chemical Individuality

The title of Garrod's landmark report on the first of his inborn errors of metabolism was "The Incidence of Alkaptonuria, a Study in Chemical Individuality."<sup>34</sup> By "incidence," he meant occurrence, and, in the work, he referred particularly to the role of parental consanguinity. In one short work, he identified consanguinity as a prime factor in the occurrence of rare recessive disorders and introduced his concept of chemical individuality.

In *Inborn Factors in Disease*, a monograph published in 1931, Garrod<sup>35</sup> generalized his concept of chemical individuality to encompass all disease, including common disorders. His thinking was rediscovered by Charles Scriver and Barton Childs, who, in 1989, published a facsimile edition of the 1931 monograph, with commentary.<sup>36</sup> They pointed out that the substance of Garrod's thesis is contained in the following summarizing paragraph at the end of his 1931 "essay"<sup>35(p 157)</sup>:

It might be claimed that what used to be spoken of as a diathesis is nothing else but chemical individuality. But to our chemical individualities are due our chemical *merits* as well as our chemical shortcomings; and it is more nearly true to say that the factors which confer upon us our predispositions to and immunities from the various mishaps which are spoken of as diseases, are inherent in our very chemical structure and even in the molecular groupings which confer upon us our individualities, and which went to the making of the chromosomes from which we sprang.

Largely on the basis of Garrodian thinking, Childs, in 1999,<sup>37</sup> developed what he called "a logic of medicine," defining logic as a statement of the formal principles underlying a branch of knowledge. From these analyses came a vision of individualized medicine—a brand of medicine designed to match the uniqueness of the individual and encompassing all disease, including common disease. Despite its title *Mendelian Inheritance in Man*, there are reasons why identifiable genetic factors in all disease including those that are not strictly Mendelian should be included in OMIM (see below). These are the common disorders previously labeled "multifactorial" and now usually termed "complex traits" (or disorders). The more we know about classic Mendelian disorders, the more we realize that these are also complex; see the example of glycerol kinase deficiency (MIM #307030).<sup>38</sup> Conversely, Mendelian subtypes of common complex disorders have come to light.

Most forms of cancer are clearly multifactorial, indeed multigenic. All are fundamentally genetic, based on changes in the genetic material; for the most part, they are somatic genetic disorders. Epigenetic changes are also importantly involved, as discussed below. In many sporadic forms of cancer, multiple genes have been identified as playing a role in initiation, progression, invasion, metastasis, and resistance to therapy. OMIM records these somatic mutations among AVs. Somatic mutations related



to prostate cancer (MIM #176807) are recorded for at least eight genes and, in the case of some of these genes, both familial and sporadic forms of prostate cancer are represented. Colorectal cancer (MIM #114500) displays an even more extensive array of genes involved in familial and/or sporadic forms.

In several instances, the gene mutant in familial cancer syndromes has been found to undergo somatic mutation to cause sporadic cancer of the type featured in the familial cancer syndrome. The *APC* gene (MIM +175100) mutant in adenomatous polyposis coli is importantly involved in sporadic colorectal cancer, and somatic mutations in *APC* have been found also in sporadic gastric cancer (MIM +175100.0010), sporadic hepatoblastoma (MIM +175100.0024), and other sporadic cancers. The *VHL* gene (MIM \*608537) is mutant in von Hippel-Lindau syndrome (MIM #193300), which has renal cancer, pheochromocytoma, and cerebellar hemangioblastoma as components; it is implicated also in sporadic cases of these three neoplasms. Germline mutations in the *TP53* gene (MIM +191170), somatic mutations of which have been identified in a variety of cancers, are the basis for one form of the Li-Fraumeni family cancer syndrome (LFS1 [MIM #151623]) that combines malignancies of a variety of tissue types, most often soft-tissue sarcomas, osteosarcomas, and breast cancer. (Li-Fraumeni syndrome is genetically heterogeneous; in addition to the LFS1 form caused by mutations in *TP53*, another form, LFS2 [MIM #609265], is caused by mutations in the *CHEK2* gene [MIM \*604373], and a third form, LFS3 [MIM %609266], maps to a locus on 1q23.)

### Extensions on Mendelism in MIM/OMIM

Classically, the determinants of variation are divided into genetic and environmental (a.k.a. exogenous or nongenetic). Random variation, "chance," is also an important determinant. A useful demonstration of both genetics and chance (stochastic variation) in determination of a particular phenotype is provided by dermatoglyphic patterns (MIM %125590): presumably, "fingerprints" are different in every human being, even identical twins. The basic differences are laid down by the DNA of the individual; additional differentiation is provided by stochastic differences in the embryologic development of the finger pads, even in individuals with a shared genome, identical twins. (For a discussion of the difference between DNA fingerprint and dermatoglyphic fingerprints and an illustrative comparison of the two types in a pair of MZ twins, see the report of the National Research Council on DNA technology in forensic science.<sup>39</sup>)

According to the role of genetic factors in pathogenesis, I and others found it useful in the early stages of the development of medical genetics to divide disease, rather arbitrarily to be sure, into Mendelian, chromosomal, and multifactorial.<sup>40</sup> The arbitrary nature of this classification does not detract from its usefulness; all classifications are

to some extent artificial. Consistent with the Garrodian perspective, all genetic variation, including that with only a contributing role in the multifactorial basis of a complex trait and perhaps even some chromosomal (genomic) variation, should be cataloged in OMIM.

In the past 20 years, epigenetic variation has become evident as a fourth major etiopathogenetic class, especially in cancer.<sup>41,42,43</sup>

### Chromosomal Variation and MIM/OMIM

Chromosomal variations (aberrations) are of interest for MIM/OMIM for several reasons. These include the role of specific genes deleted or duplicated in aneuploid states in determining phenotype; see Down syndrome (MIM #190685), Wolf-Hirschhorn syndrome (MIM #194190), cri-du-chat syndrome (MIM #123450), Emanuel syndrome (MIM #609029), 22q11.2 deletion syndrome (MIM #188400), Jacobsen syndrome (MIM #147791), 9q subtelomeric deletion syndrome (MIM #610253), and others.

A large interest of MIM/OMIM in chromosomal aberrations is in connection with gene mapping. Notable examples of identification of the chromosomal locus of Mendelian disorders through finding small interstitial deletions in sporadic cases include retinoblastoma (MIM +180200) on chromosome 13, Duchenne muscular dystrophy on Xp, and adenomatous polyposis coli on chromosome 5. In all three cases, it was the cytogenetic clue that led to isolation of the mutant gene. One of the earliest examples of deletion mapping was assignment of the ABO blood group—adenylate kinase—nail-patella syndrome cluster of gene loci to 9q34 by Ferguson-Smith et al.<sup>44</sup> Aniridia (MIM #106210) was mapped to 11p13 by its occurrence alone or as part of the Wilms-aniridia-genitourinary-mental retardation syndrome (WAGR [MIM #194072]) in patients with deletions in that region of 11p.

Deletions have also been a clue to linkage of loci. Schmickel's concept of contiguous gene-deletion syndromes<sup>45</sup> turned on its ear the idea that syndromes are always based on pleiotropism of a single mutant gene and are never due to close linkage of two or more genes, each of which is responsible for individual components of the syndrome. Notable examples of contiguous gene-deletion syndromes in addition to the just-mentioned WAGR include Langer-Giedion syndrome (MIM #150230), Miller-Dieker lissencephaly syndrome (MIM #247200), and Williams syndrome (MIM #194050). Molecular cytogenetics has aided greatly in the elucidation of this topic. In each of the examples cited, specific genes that are deleted have been identified.

Reciprocal X-autosome translocations have been important for finding the chromosomal location of genes determining X-linked disorders (and the nature of those genes). Sporadic cases of an X-linked recessive disorder in a female with such a translocation was sometimes not only the main or even the only evidence that the gene was on the X chromosome, but also the breakpoint on X indicated

the precise location of the gene. In females with an X-autosome translocation, the derivative X chromosome (the derivative chromosome with the centromere of the X) is active in all cells; if the derivative X chromosome were inactive (lyonized), its autosomal component would likewise be inactive, with cell-lethal effects due to autosomal monosomy. Thus, the normal X chromosome is the inactive one, and the derivative X—which may have disruption of the gene at the breakpoint—is the active one. The breakpoint, therefore, marks the site of the gene responsible for the disorder in the patient. Duchenne muscular dystrophy was among the first disorders to be mapped by this method; table 1 in the preface of MIM12 (1998) tabulated 16 other examples.<sup>2</sup>

De novo reciprocal autosomal translocations also can provide information on the chromosomal site of autosomal Mendelian disorders or specific autosomal genes. Single cases of de novo autosome-autosome translocation are less informative than are single cases of X-autosome translocation, because either autosome can be the site of the gene disrupted by the chromosome break, or perhaps the phenotype may be the result of a fusion gene—for example, a joining of the promoter region of one gene and the coding region of the other. The occurrence of two or more reciprocal translocations that involve the same chromosome as one of the partners and involve the same chromosome band establishes the site of the gene of interest. For example, in the case of type II (ankyrin-related) hereditary spherocytosis (MIM +182900), the disorder was mapped to 8p by the discovery of a family with an apparently balanced translocation between chromosomes 8 and 12<sup>46,47</sup> and another between chromosomes 3 and 8<sup>48</sup>; in each family, spherocytosis segregated with the balanced translocation, and the break in 8p was at the same site.

Many sporadic reciprocal translocations of specific types have been found in particular hematologic malignancies; >25 were listed in table 3 of the preface of MIM12 (1998), and many more have been described since then<sup>49,50</sup>; far fewer reciprocal translocations have been found in solid tumors. All are sarcomas: Ewing sarcoma results from fusion of the *EWS* gene (MIM +133450) on chromosome 22 with genes elsewhere in the genome, such as *FLI1* (MIM \*193067) on chromosome 11 in the translocation t(11;22)(q23;q12). Other examples are synovial sarcoma (*SSX1* [MIM +312820] and *SS18* [MIM \*600192]) and myxoid liposarcoma (*FUS* [MIM \*137070] and *CHOP* [MIM +126337]), resulting from translocations t(X;18)(11.2;q11.2) and t(12;16)(q13.3;p11), respectively. The *PAX3* gene (MIM \*606597), which is mutant in Waardenburg syndrome (MIM #193500), when fused with the *FKHR* gene (MIM \*136533) by translocation t(2;13)(q35;q14), gives rise to alveolar rhabdomyosarcoma (MIM #268220).

Molecular genetics study of the translocations in hematologic malignancies, with characterization of the genes that are fused or disrupted by the process, has contributed information for creation of many new gene en-

tries in OMIM. The record for promiscuity probably goes to the gene called “*MLL*” (MIM +159555) for “myeloid/lymphoid leukemia.” Situated on 11q23, the *MLL* gene partners with genes at 15 or more other chromosomal sites in reciprocal translocation, to result in mixed-lineage type leukemia—for example, with *AF4* (MIM \*159557) in t(4;11)(q21;q23), *AF6* (MIM #159551) in t(6;11)(q27;q23), *AF9* (MIM \*159558) in t(9;11)(p22;q23), and *ENL* (MIM \*159556) in t(11;19)(q23;p13).

Inversions can also cause birth defects or neoplasia, either by gene disruption produced by the break at one end or by the bringing together of control elements of one gene with the coding portion of another. An inversion in the long arm of chromosome 2 was a clue to the location of the gene for Waardenburg syndrome type 1. Inversions and other aberrations involving 16p were the only evidence of the location of the gene mutant in Rubinstein-Taybi syndrome (MIM #180849).

Several Mendelian disorders are characterized by increased chromosomal breakage occurring spontaneously or induced by clastogenic agents and appear to be accompanied by a predisposition to malignancy. These chromosomal breakage syndromes include Fanconi anemia (MIM #227650), Bloom syndrome (MIM #210900), ataxia-telangiectasia (MIM #208900), Nijmegen breakage syndrome (MIM #251260), LIG4 syndrome (MIM #606593), and ICF syndrome (MIM #242860); the molecular defect is known in each of these disorders. Some of the most striking chromosomal changes, referred to as “heterochromatic splaying” or “heterochromatic repulsion,” occur in Roberts syndrome (MIM #268300). This and the SC phocomelia syndrome (MIM #269000) are caused by mutations in the *ESCO2* gene (MIM \*609353), whose gene product is required for establishment of sister chromatid cohesion during the S phase of the cell cycle.

Molecular cytogenetics and molecular genetics in general have narrowed the gap between Mendelian genetics and the classic cytogenetics of clinical disorders. This is illustrated by the conditions termed “genomic disorders” by Lupski,<sup>51</sup> many of which show Mendelian patterns of inheritance.

Variation in gene copy number, through gains (duplications) or losses (deletions) of chromosome segments, is as extensive and as potentially significant in disease as that represented by SNPs.<sup>52</sup> Variation in gene copy number has been demonstrated in many neoplasms, including cancer of breast, prostate, ovary, colon, head and neck, brain, and pancreas,<sup>53</sup> as well as lymphoma and adenocortical cancer. It has been suggested<sup>54</sup> that the extraordinary copy-number polymorphism of the complement component-4 (*C4A* [MIM +120810]) may be related to defense against infectious disease and susceptibility to autoimmune disease. Similarly, it is thought<sup>55</sup> that an increased copy number of the trypsinogen-1 gene (*PRSS1* [MIM +276000]) may account for some of the families with hereditary pancreatitis without a known causative mutation in that gene. The beta-defensin genes (e.g., MIM \*602215), clustered on

8p2.1, vary in copy number from 2 to 12. Low copy number is associated with susceptibility to Crohn disease (MIM #266600). The *CCL3L1* gene (MIM \*601395), which occurs in variable copy number (1–10), shows an association of copy number with susceptibility to HIV/AIDS (MIM #609423).<sup>56</sup> Copy-number variation of the orthologous rat and human *FCGR3B* gene (MIM \*610665) is a determinant of susceptibility to immunologically mediated glomerulonephritis; low copy number was found to be associated with glomerulonephritis in systemic lupus erythematosus (MIM #152700).<sup>57</sup>

Gene duplication is a well-known basis of Mendelian disorders. Most cases of Pelizaeus-Merzbacher disease (MIM #312080) are caused by duplication of the *PLP1* gene (MIM \*300401). In many instances, Charcot-Marie-Tooth disease type 1a (MIM #118210), another disorder of myelination, is caused by duplication of the gene encoding peripheral myelin protein-22 (*PMP22* [MIM \*601097]). Gene duplication is sometimes the cause of disease in a small but nonetheless instructive subsets of patients, including patients with CHARGE syndrome (MIM #214800) caused by duplication in the *CHD7* gene (MIM \*608892),<sup>58</sup> patients with Parkinson disease (MIM #168601) caused by duplication in the *SNCA* gene (MIM \*163890.0005), and patients with a form of early-onset Alzheimer disease with cerebral amyloid angiopathy (MIM #104300) caused by duplication in the *APP* gene (MIM \*104760.0020).

Twenty-three chromosomal fragile sites are described in OMIM. Lubs<sup>59</sup> discovered the first fragile site, FRAXA, on Xq (*FMR1* [MIM \*309550]). FRAXA is associated with fragile X mental retardation syndrome (MIM #300624), the most common genetic cause of mental retardation after Down syndrome. A fragile site on chromosome 16q22 (*FRA16A* [MIM #136580]) is of historical interest, since it was the chromosome marker used to confirm the assignment of the haptoglobin locus (MIM \*136580) to chromosome 16 by studies in somatic cell hybrids.<sup>60</sup>

## Multifactorial Disorders (“Complex Traits”) and MIM/OMIM

Traits in which variation in multiple loci/genes collaborate with multiple nongenetic (“environmental”) factors, so-called complex traits, also have their place in MIM/OMIM, but the way these phenotypes and the underlying genetic factors are recorded requires special consideration. The complex traits include all common disorders such as essential hypertension, mental illness, asthma, and so many more. Some have rare Mendelian subtypes that represent no problem for entry in OMIM: Lifton and his colleagues<sup>61,62</sup> tabulated eight rare Mendelian forms of hypertension, each due to a disturbance in the handling of sodium by the renal tubules.

The demonstration by association or linkage studies of a relationship between specific variation in a specific gene and a specific complex disorder has become prominent in recent years. With these methods, for example, Chang et

al.<sup>63</sup> described three blood pressure–related genes on chromosome 1q42. In that study, individual variants in these three genes accounted for differences of 2–5-mm Hg in mean systolic blood pressure levels, and the cumulative effect reached 8–10 mm Hg. Such information on susceptibility (or resistance) genes has been recorded in OMIM in the entry for the gene (as well as in the entry for the phenotype) and is represented by an AV under the gene entry.

Associations with a particular haplotype are recorded under the phenotype entry if a relationship to a specific gene is not clear. Some QTLs are recorded as separate entries in MIM/OMIM, their existence demonstrated by linkage and association studies. These include QTLs for obesity (e.g., MIM %602025), stature (e.g., MIM %606255), level of high-density lipoprotein (e.g., MIM %606613), bone density (e.g., MIM %601884), intelligence (e.g., MIM %603783), hemoglobin level (e.g., MIM %609319), and mean telomere length (e.g., MIM %609113).

## Oligogenic Inheritance

OMIM describes >14 instances of possible digenic inheritance. The first molecular documentation of digenic inheritance, reported by Kajiwara et al.,<sup>64</sup> concerned retinitis pigmentosa caused by a heterozygous mutation in the gene encoding peripherin (*RDS* [MIM \*179605.0004]) in combination with a heterozygous null mutation in the unlinked gene (*ROM1* [MIM \*180721.0001]) encoding rod outer segment protein-1. Retinitis pigmentosa did not occur with either mutation alone in the heterozygous state. Nadeau<sup>65</sup> suggested that this is an example of modification rather than digenic inheritance, *ROM1* being the modifier. He cited classic examples of dominance modification in mouse models.

Katsanis et al.<sup>66</sup> reported several families in which Bardet-Biedl syndrome (BBS [MIM #209900]) showed what they termed “triallelic inheritance”—for example, patients with BBS who are homozygous for a missense mutation in the *MKKS* gene (MIM \*604896.0003) and heterozygous for a mutation in the *BBS2* gene (MIM \*606151.0013). Katsanis et al.<sup>66</sup> estimated that 40% of patients with BBS-2 (see MIM #209900) have homozygosity or compound heterozygosity for mutations in the *BBS2* gene in combination with heterozygosity for a third mutation in another BBS gene.

Bartter syndrome type 4 (MIM #602522)—renal salt wasting and deafness—is most often caused by mutation in the *BSND* gene (MIM \*606412). In a child with this disorder whose parents were consanguineous, Schlingmann et al.<sup>67</sup> found no mutation in the *BSND* gene but found homozygous deletion of the *CLCNKB* gene (MIM \*602023.0008) and a homozygous missense mutation of the linked *CLCNKA* gene (MIM \*602024.0001).

In some cases, the nature of the interaction of the gene products in a triallelic digenic inheritance pattern can be deduced—for example, in the case of cortisone reductase deficiency (MIM #604931) due to an intronic mutation in

*HSD11B1* (MIM \*600713.0001) and exonic mutations in *H6PD* (MIM \*138090.0001–\*138090.0002).<sup>68</sup> The long QT syndromes, in which the interaction of mutations in two different ion-channel genes appear to occur, provide other examples. OMIM records two examples of the LQT syndrome resulting from double heterozygosity for mutations in two LQT genes—that is, biallelic digenic inheritance. LQT 1/2 results from heterozygous mutations in the *KCNQ1* (MIM \*607542.0009) and *KCNH2* (MIM +152427.0019) genes, and LQT3/6 results from heterozygous mutations in the *SCN5A* (MIM +600163.0007) and *KCNE2* (MIM +603796.0005) genes. LQT 2/5 can result from a heterozygous mutation in the *KCNH2* (MIM +152427.0021) gene and a homozygous mutation in the *KCNE1* (MIM \*176261.0005) gene, another example of triallelic digenic inheritance.

As more is learned about Mendelian disorders, complexities come to light that indicate that most of these also must be viewed as multifactorial or at least as complex traits.<sup>38</sup> Modifier genes/loci implicated in Mendelian disorders are being identified—for example, the cystic fibrosis modifier-1 locus (*CFM1* [MIM 603855]) in cystic fibrosis. Imprinting is an important factor contributing to the complexity of inheritance in a number of genetic disorders. In the instance of the many Mendelian disorders that are caused by expanded repeats, the random loss and gain in number of repeats introduce complexities.

### Susceptibility Alleles

Susceptibility alleles represent a major part of the OMIM record of the multifactorial basis of common disorders. In the case of ~490 of the ~2,000 “disease genes” (genes with one or more disease-related AVs), at least one of the AVs is a susceptibility (or resistance, protection) allele. Of the 3,345 phenotypes for which the molecular basis is established, 375 (11%) are susceptibility phenotypes. By use of braces or the words “susceptibility to,” these 375 phenotypes are identified as predisposed to or protected against in the “Disorder” field of the OMIM Gene Map; a specific molecular basis is indicated by an appended “(3).”

Many of the susceptibility alleles are common polymorphisms, by convention defined as “variations with an allele frequency of >.01 (1%).” The relationship of the polymorphism to the specific disorder has been identified mainly by association, linkage, and transmission/disequilibrium studies. One of the earliest susceptibility alleles to be identified was that related to Alzheimer disease (MIM #104310), the APOE4 variant of apolipoprotein E (Cys112Arg [MIM +107741.0016]).<sup>69,70</sup>

In many cases, more than one common disorder is related to the same polymorphic allele. In such cases, which common disorder results is thought to depend on the rest of the genetic constitution of the individual and particular environmental circumstances. The different disorders related to a particular susceptibility allele sometimes have obvious possible connections, as in the case of different forms of autoimmune disorders. Even in different indi-

viduals in the same family, the predisposition conveyed by the susceptibility allele may take the form of different common disorders. These considerations are the basis of Becker’s common alleles/multiple common disorders model of the genetics of common disease.<sup>71</sup>

Several different types of autoimmune disease (MIM #109100) have been found to be associated with particular susceptibility alleles, even in the same family. For example, the Arg620Trp polymorphism of the *PTPN22* gene (MIM \*600716.0001) has been identified as a susceptibility allele in insulin-dependent diabetes mellitus, rheumatoid arthritis, systemic lupus erythematosus, and Hashimoto thyroiditis. The insertion/deletion polymorphism of angiotensin I-converting enzyme (ACE [MIM +106180.0001]) has been identified as a susceptibility allele for myocardial infarction, diabetic nephropathy, hemorrhagic stroke, ischemic stroke, and progression of severe acute respiratory syndrome. In addition, the indel polymorphism of ACE appears to be a QTL for stature. The Val66Met polymorphism of the gene encoding brain-derived neurotrophic factor (MIM \*113505.0002) has been related to susceptibility to memory impairment, anorexia nervosa, bulimia nervosa, and bipolar affective disorder, as well as to protection against obsessive-compulsive disorder and modification of the age at onset of Parkinson disease.

The common disorders to which susceptibility alleles have been related include obesity (alleles in 9 different genes), SLE (in 4), osteoporosis (in 4), osteoarthritis (in 4), asthma (in 11), myocardial infarction (in 13), coronary artery disease (5 susceptibility alleles and 1 resistance allele), and hypertension including preeclampsia (in 13). Susceptibility or resistance alleles have also been identified for a variety of infections, including tuberculosis, leprosy, *Helicobacter pylori*, Legionnaire disease, cerebral malaria, and, notably, HIV/AIDS. Neuropsychiatric disorders with identified susceptibility alleles include schizophrenia (in five genes), various forms of affective disorder (in four), autism (in five), and obsessive-compulsive disorder (in two).

### Epigenetic Variation in MIM/OMIM

In both the phenotype and the gene entries in MIM/OMIM, much information is recorded about epigenetic variation. This is variation in gene expression that is not encoded in the DNA sequence itself. The many areas of epigenetics touched on in MIM/OMIM include aspects of X-chromosome inactivation, autosomal imprinting, the role of methylation and histone modification, and the implications for the pattern of inheritance and phenotype of Mendelian disorders and for the pathogenesis of certain developmental abnormalities such as Beckwith-Wiedemann syndrome (MIM #130650) and many cancers.

Discussions of X-chromosome inactivation include the role of specific genes in the creation or maintenance of the inactive state—for example, *XIST* (MIM \*314670),

TSIX (MIM \*300181), and XCE (MIM \*300074). Familial skewed X inactivation (MIM #300087) is sometimes due to mutations in the *XIST* gene; skewed X inactivation occurs also in women heterozygous for some X-linked disorders, including X-linked severe combined immunodeficiency (MIM #300400), Wiskott-Aldrich syndrome (MIM #301000), and dyskeratosis congenita (MIM #305000), in which cells with the mutation-carrying X chromosomes are at a selective disadvantage. In such instances, the skewing can be used as a method for diagnosing the heterozygous carrier state. In some instances, skewed inactivation may be a main piece of evidence establishing X-linked inheritance; for example, Aicardi syndrome (MIM #304050) is probably X-linked dominant with male lethality, to judge from the finding of skewed X inactivation in females.

Autosomal imprinting is involved in the phenotypic consequences of uniparental disomy. A prime example is provided by Prader-Willi syndrome (MIM #176270) in individuals with maternal uniparental disomy for chromosome 15; since both chromosomes 15 come from the mother, the individual lacks the paternally expressed gene *IPW* (MIM \*601491) located in proximal 15q.

The epigenetic silencing of tumor-suppressor genes may be a more frequent basis for cancer than are point mutations in those genes.<sup>43</sup> Aberrant promoter methylation is associated with loss-of-gene function that can provide a selective advantage to neoplastic cells, just as do loss-of-function point mutations. Germline mutations in the *VHL* (MIM \*608537), *BRCA1* (MIM +113705), and *STK11* (MIM \*602216) genes cause familial forms of renal, breast, and colon cancers, respectively; the same genes are often epigenetically silenced in sporadic forms of these tumors. For example, the *BRCA1* gene is not important only for familial breast cancer; 10%–15% of women with nonfamilial breast cancer have tumors in which the *BRCA1* gene is hypermethylated.

Thus, in addition to the >2,000 genes for which one or more specific disease-related mutations have been found, other genes cataloged in OMIM are important to the pathogenesis of cancers through epigenetic mechanisms. These are tumor-suppressor genes silenced through hypermethylation. Examples include the *RASSF1* gene (MIM \*605082) on 3p21, which is often deleted or its promoter hypermethylated in lung cancer. This gene shows anomalous promoter hypermethylation in a large number of other tumor types<sup>72</sup> as well. As indicated by their names, other tumor-suppressor genes related to cancers through epigenetic silencing are “hypermethylated in cancer-1” (*HIC1* [MIM \*603825]) on 17p and “hypermethylated in cancer-2” (*HIC2* [MIM \*607712]) on 22q.

The converse situation, activation of oncogenes through hypomethylation, also leads to the development of cancers. Indeed, hypomethylation was the first indication of the role of anomalous promoter methylation in carcinogenesis, as reported by Feinberg and Vogelstein in 1983.<sup>73</sup>

Loss of imprinting (LOI), an epigenetic alteration, has

also been found in cancers. LOI of the gene encoding insulinlike growth factor II (*IGF2* [MIM +147470]) has been described for Wilms tumor (MIM #194070), in Beckwith-Wiedemann syndrome (MIM #130650), and in hepatoblastoma. LOI of *IGF2* is found in normal colonic mucosa of ~30% of patients with colorectal cancer (MIM #114500) but in only ~10% of normal individuals. In a study of 172 patients in a colonoscopy clinic, Cui et al.<sup>74</sup> found that the adjusted odds ratio for LOI of *IGF2* in lymphocytes was 5.15 for patients with a positive family history, 3.46 for patients with adenomas, and 21.7 for patients with colorectal cancer. Other work supported the idea that LOI of *IGF2* may be a familial characteristic.

## MIM/OMIM as a Historical Document

At the outset in the 1960s, each entry in MIM was assembled and organized like those in the *Oxford English Dictionary* (OED)<sup>75,76</sup>; that is, on a chronologic, diachronic (“through time”), or historical basis, rather than the hierarchic, descriptive, or synchronic (“one point in time”) method used by textbooks and encyclopedias. After the initial creation of individual entries, information provided by new publications was usually added at the end of the existing record. Even when many entries in MIM reached a volume requiring reorganization into topical sections, the diachronic approach was maintained in each section; thus, the historical development of human genetics and particularly medical genetics is evident. The line-up of the 12 print editions of MIM (1966–1998) reflects that progress (fig. 1). The 12 editions are serial cross-sections of the field over the 33 years, 1966–1998, spaced 3 years apart, on average.

Progress in human genetics during the 40 years of the existence of MIM/OMIM can be gauged by several other measures more meaningful than the size of the books, including the total number of entries and the number of entries of particular types. Perhaps the most meaningful measures of progress are those that relate to the number of loci/genes mapped to specific chromosomal sites, the number of genes that have been identified as the site of disease-related mutations, and the total number of phenotypes that have been “molecularized”—that is, found to be associated with specific DNA mutations. Data on these three aspects of the scientometrics of human genetics are presented in graphs published elsewhere<sup>29</sup> and in tables 1–3.

In addition to the title and author indices, each of the 12 print editions of MIM contained a foreword/preface comprising a long essay on Mendelian inheritance, a description of the methods used in assembling the catalogs, and a tabulation of statistics on the journals that were sources of information and on the growth of the catalogs, in terms of total number of entries. This “front material” of the book also included a synopsis of the human gene map and a listing of molecular defects in genetic disorders. The first synopsis of the human gene map, published in

**Table 3. Molecular Defects in Mendelian Disorders (and Somatic Mutations in Neoplasms) (January 29, 2007)**

Mapping Type	No. Mapped
Loci in OMIM with at least one known point mutation that causes a disorder or neoplasm	2,002
Mapped disorders for which a causative mutation has been identified	3,345 <sup>a</sup>
Total no. of mutations cataloged in OMIM	14,949

<sup>a</sup> Number of phenotypes labeled with "(3)" in the Disorder field of the Synopsis of the Human Gene Map.

the 1971 (3rd) edition, occupied only 1 page; it expanded to 116 pages in the 1998 (12th) edition. Today, it would consume at least 256 print pages if presented in the same format as that used in the 1998 edition.

The Synopsis of the Human Gene Map is available as an appendix to OMIM on the World Wide Web. It lists, chromosome by chromosome in tabular form, the genes mapped to particular sites, beginning at the end of the short arm. The focus of this map is on the "morbid anatomy of the human genome" and is sometimes referred to as the "Morbid Map." It is here that relationships between gene and phenotype are most easily seen. The Disorder field in the tabular synopsis of the human gene map lists the different phenotypes that are due to allelic mutations in the given gene. An alphabetic listing by disorder, also available on the Web, shows *genetic (locus) heterogeneity*, or how one phenotype can have its basis in mutation in any one of several different genes/loci. The listing of the genes/loci by chromosome and the multiple distinct pathologic phenotypes that are related to mutation in a single gene best shows *phenotypic heterogeneity* at a locus.

MIM/OMIM has regularly recorded information on animal models of human disease. The tabular synopsis of the gene map includes a field for homologous mouse loci, and, starting with the 8th edition of MIM (1988),<sup>1</sup> the book edition included the Oxford Grid (e.g., p. cclxi in MIM12).<sup>2</sup> This pictorial representation of human and mouse homologies was developed by John H. Edwards, Professor of Genetics at Oxford and Harwell colleagues C. V. Buckle, A. G. Searle, V. J. Buckle, and others.<sup>77,78,79</sup> On the principle of homology of synteny, the information in the Oxford Grid often provided an initial clue to the location of the homologous gene in the human or provided support for the chromosomal location identified in the human. With OMIM on the Web, links can now be made to several model organism resources, including the Mouse Genome Database (Mouse Genome Informatics) at The Jackson Laboratory and Online Mendelian Inheritance in Animals (OMIA),<sup>80</sup> through the Entrez Search and Retrieval System of the NCBI.

### Use of OMIM in the Clinic and in Research and Teaching

As a comprehensive account of the state of knowledge of the genetic basis of health and disease, OMIM is intended to have wide usefulness to researchers, clinicians, and stu-

dents. OMIM's aid to the clinician comes particularly in the area of differential diagnosis and other aspects of the large number of individually rare Mendelian disorders to which the human is literally heir. Precise diagnosis is the basis for quality care and for accurate genetics counseling and appropriate therapy. It appears that many clinicians and counselors use OMIM as a reference in patient care. Information on prognosis, potential complications, and possibilities for prenatal diagnosis and carrier detection is important to the care of patients, both by specialists in medical genetics and by other health professionals who may rarely encounter these disorders. The diagnostic usefulness of OMIM is enhanced by the Clinical Synopsis section provided with most phenotype entries. Through NCBI, OMIM is linked to other online clinical resources such as the highly useful GeneTests, a directory of >600 laboratories that perform diagnostic testing on >1,300 disorders (November 2006). GeneTests also includes a large number of frequently revised reviews of various genetic disorders, prepared by clinical experts in a standardized synchronic form, as an aid to clinical and laboratory diagnosis.

The usefulness of OMIM extends beyond the clinic. Molecular biologists find OMIM useful for background information on genes and descriptions of disorders in their particular areas of research interest. OMIM lends itself well both to research "mining" and to educational "surfing." I have always thought that a great deal can be learned (and taught) about human genetics, clinical genetics, and human molecular genetics by means of exercises consisting of questions to which the answers can be found in OMIM. In 1993, I produced a self-instruction guide and workbook for use with OMIM.<sup>81</sup> An interactive online workbook might be particularly useful.

### Future Challenges of OMIM

As outlined earlier, an objective of OMIM from its beginning as the book MIM has been to catalog the relationship between phenotype and gene. The organization of MIM/OMIM has evolved to accommodate that fundamental objective of human genetics. Mapping has been important in establishing that relationship and has been comprehensively chronicled in MIM/OMIM. With completion of the sequencing of the human genome, all genes have, in effect, been mapped in terms of their location in the sequence, but many remain to be characterized ("anno-

tated"). The goal now is to relate phenotype to gene function.

Despite the official completion of the human genome project several years ago, the number of identified genes continues to rise, although more slowly. Extrapolating the growth of the number of genes with "known or inferred function" suggests that, barring major surprises (such as the growing number of small RNAs), nearly all genes will belong to this category in the next 6–8 years. Of course, "known" is a relative term; we continue to learn new biology about a great many of the genes already in OMIM. A challenge for OMIM is to capture information describing "new" genes while continuing to add important science to the ~11,500 current gene entries for which the sequence is known. Other challenges will be to capture alias gene names and to register in some orderly way the complexity of combinatorial alternative splicing and many other surprises that the human genome is likely to reveal.

Mendelian (monogenic or monolocus) phenotypes and their molecular basis will continue to be the principal fodder for the ever enlarging catalogs.<sup>82</sup> The mechanism by which the mutation leads to the phenotype (the steps from gene to phenotype) will always be important information for cataloging. There will be attention also to collation of information on genetic, epigenetic, and environmental modifiers of Mendelian phenotypes. A challenge OMIM already faces is how to catalog complex phenotypes and complex genotypes and their functional relationships to each other and to include epigenetics (and epigenomics), the interaction of genes and gene products, the interaction with and influence of environment, and the emergent phenotypes resulting from these interactions—no small undertaking.

These relationships between complex phenotypes and complex genotypes are under investigation in a large number of clinical and epidemiologic research programs, usually involving, by necessity, large cohorts of subjects, with use of haplotype data for description of the genotype in association studies, and covering a range of topics for study that includes cancer(s), cardiovascular disease(s), asthma, and mental illnesses. OMIM must continue to register this information in a manner that is useful to clinical medicine and that promotes our fundamental understanding of the genetics of health and disease.

Will there ever be another print edition of MIM? At this time, that seems unlikely because of the obvious advantages of the electronic version, with its daily updating and its searchability. The availability of the book in nonelectronic settings and the tables and other appendiceal material in the preface and foreword may not justify its existence. The historian in me regrets the loss of the archival function of the print edition.

My colleague Alan Scott argues that there would be a use for a print version of an annotated human genome atlas with abbreviated MIM entries. He points out that a book is easier to browse, with opportunities for serendipity to operate, than is a computerized database. Such an atlas

could be organized chromosome by chromosome with, in effect, a separate "catalog" for each autosome comparable to the present chromosome-specific catalogs (X, Y, and mitochondrial).

## Acknowledgments

Dozens of colleagues at Johns Hopkins and elsewhere have contributed to the creation and maintenance of MIM and OMIM over the past 45 years. For several years, the curation of OMIM at Johns Hopkins has been supported by a contract from the National Library of Medicine funded by the National Human Genome Research Institute. Ada Hamosh, M.D., is scientific director of the OMIM project; Alan F. Scott, Ph.D., is deputy director for genes; Joanna S. Amberger is project manager; and Carol A. Bocchini is senior editor and writer.

## Web Resources

The URLs for data presented herein are as follows:

Cystic Fibrosis Mutation Database, <http://www.genet.sickkids.on.ca/cftr/app>  
Entrez, <http://www.ncbi.nlm.nih.gov/entrez/>  
GeneTests, <http://www.genetests.org/>  
HUGO, <http://www.gene.ucl.ac.uk/nomenclature/>  
Human Gene Mutation Database, <http://www.hgmd.cf.ac.uk>  
Human Genome Variation Society, <http://www.genomic.unimelb.edu.au/mdi/dblist/glsdb.html>  
Mouse Genome Informatics, <http://www.informatics.jax.org/> (for the Mouse Genome Database)  
Online Mendelian Inheritance in Animals (OMIA), <http://omia.angis.org.au/>  
Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/>  
PAHdb, <http://www.pahdb.mcgill.ca/>

## References

References to many other publications related to particular disorders and genes reviewed here are provided in specific gene or phenotype entries in OMIM (identified by their 6-digit entry numbers).

1. McKusick VA. Mendelian inheritance in man: catalogs of autosomal dominant, autosomal recessive and X-linked phenotypes (1966, 1st ed; 1969, 2nd ed; 1971, 3rd ed; 1975, 4th ed; 1978, 5th ed; 1983, 6th ed; 1986, 7th ed; 1988, 8th ed; 1990, 9th ed; 1992, 10th ed). Johns Hopkins University Press, Baltimore
2. McKusick VA. Mendelian inheritance in man: a catalog of human genes and genetic disorders (1994, 11th ed; 1998, 12th ed). Johns Hopkins University Press, Baltimore
3. McKusick VA (2006) A 60-year tale of spots, maps, and genes. *Annu Rev Genomics Hum Genet* 7:1–27
4. McKusick VA and contributors (1961) *Medical Genetics 1958–1960: an annotated review*. Mosby, St. Louis
5. McKusick VA and contributors (1965) *Medical genetics 1961–1963: an annotated review*. Pergamon Press, Oxford, United Kingdom
6. McKusick VA (1962) On the X chromosome of man. *Quart Rev Biol* 37:69–175
7. McKusick VA (1964) On the X chromosome of man. *American Institute of Biological Sciences, Washington, DC*

8. McKusick VA, Hostetler JA, Egeland JA (1964) Genetic studies of the Amish: background and potentialities. *Bull Johns Hopkins Hosp* 115:203–222
9. McKusick VA, Hostetler JA, Egeland JA, Eldridge R (1964) The distribution of certain genes in the Old Order Amish. *Cold Spring Harb Symp Quant Biol* 29:99–113
10. Antonarakis SE and the Nomenclature Working Group (1998) Recommendations for a nomenclature system for human gene mutations. *Hum Mutat* 11:1–3
11. Den Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15:7–12
12. Den Dunnen JT, Antonarakis SE (2001) Nomenclature for the description of human sequence variations. *Hum Genet* 109:121–124
13. Migeon B, Miller CS (1968) Human-mouse somatic cell hybrids with single human chromosome (group E): link with thymidine kinase activity. *Science* 162:1005–1006
14. Miller OJ, Allderdice PW, Miller DA (1971) Human thymidine kinase gene locus: assignment to chromosome 17 in a hybrid of man and mouse cells. *Science* 173:244–245
15. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33:D514–D517
16. Harman D, Benson D, Fitzpatrick L, Huntzinger R, Goldstein C (1988) IRx: an information retrieval system for experimentation and user applications. *Proceedings of RIAO 88 Conference "User-Oriented Content-Base Text and Image Handling,"* Cambridge, MA, March 21–24
17. Faber KH (1930) *Nosography*, 2nd ed. Paul B Hoeber, New York
18. Bergsma D, McKusick VA, Dorst JP, Scott CI (eds) (1969) The first conference on the clinical delineation of birth defects. Part IV. Skeletal dysplasias. *Birth Defects Orig Art Ser* 5:1–396
19. McKusick VA (1969) Lumpers and splitters, or the nosology of genetic disease. *Birth Defects Orig Art Ser* 5:23–32
20. McKusick VA (1969) Lumpers and splitters, or the nosology of genetic disease. *Perspect Biol Med* 12:298–312
21. McKusick VA (1956) *Heritable disorders of connective tissue*, 1st ed. Mosby, St. Louis
22. Feingold M (2006) The use of inappropriate, demeaning, and pejorative terminology to describe syndromes. *Am J Med Genet A* 140:410–411
23. Cohen MM Jr (2006) Problems in the naming of genes. *Am J Med Genet A* 140:1483–1484
24. Ludman MD (2006) The use of inappropriate, demeaning, and pejorative terminology in gene nomenclature: a comment on Feingold. *Am J Med Genet A* 140:1485–1486
25. Wain HM, Lovering RC, Bruford EA, Lush MJ, Wright MW, Povey S (2002) Guidelines for human gene nomenclature. *Genomics* 79:464–470
26. Goodfellow P (1997) A celebration and a farewell. *Nat Genet* 16:209–210
27. Donahue RP, Bias WB, Renwick JH, McKusick VA (1968) Probable assignment of the Duffy blood group locus to chromosome 1 in man. *Proc Nat Acad Sci USA* 61:949–955
28. Antonarakis SE, McKusick VA (2000) OMIM passes the 1,000-disease-gene mark. *Nat Genet* 25:11
29. Peltonen L, McKusick VA (2001) Dissecting human disease in the postgenomic era. *Science* 291:1224–1229
30. Slaney SF, Oldridge M, Hurst JA, Morriss-Kay GM, Hall CM, Poole MD, Wilkie AOM (1996) Differential effects of FGFR2 mutations on syndactyly and cleft palate in Apert syndrome. *Am J Hum Genet* 58:923–932ha
31. Garrod AE (1928) The lessons of rare maladies. *Lancet* 1:1055–1060
32. Gorlin JB, Yamin R, Egan S, Stewart M, Stossel TP, Kwiatkowski DJ, Hartwig JH (1990) Human endothelial actin-binding protein (ABP-280, nonmuscle filamin): a molecular leaf spring. *J Cell Biol* 111:1089–1105
33. Gorlin JB, Henske E, Warren ST, Kunst CB, D'Urso M, Palmieri G, Hartwig JH, Bruns G, Kwiatkowski DJ (1993) Actin-binding protein (ABP-280) filamin gene (FLN) maps telomeric to the color vision locus (R/GCP) and centromeric to G6PD in Xq28. *Genomics* 17:496–498
34. Garrod AE (1902) The incidence of alkaptonuria: a study in chemical individuality. *Lancet* 2:1616–1620
35. Garrod AE (1931) *The inborn factors in disease: an essay*. Clarendon Press, Oxford, United Kingdom
36. Scriver CR, Childs B (1989) *Garrod's inborn factors in disease*. Oxford University Press, New York
37. Childs B (1999) *Genetic medicine: a logic of disease*. Johns Hopkins University Press, Baltimore
38. Dipple KM, Zhang YH, Huang BL, McCabe LL, Dallongeville J, Inokuchi T, Kimura M, Marx HJ, Roederer GO, Shih V, et al (2001) Glycerol kinase deficiency: evidence for complexity in a single gene disorder. *Hum Genet* 109:55–62
39. National Research Council (1992) *DNA technology in forensic science*. National Academy Press, Washington, DC
40. McKusick VA (1964) Lewis A Conner memorial lecture: a genetical view of cardiovascular disease. *Circulation* 30:326–357
41. Feinberg AP, Tycko B (2004) The history of cancer epigenetics. *Nat Rev Cancer* 4:143–153
42. Egger G, Ling G, Aparicio A, Jones PA (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429:457–463
43. Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Cancer* 3:414–428
44. Ferguson-Smith MA, Aitken DA, Turleau C, de Grouchy J (1976) Localisation of the human ABO: Np-1: AK-1 linkage group by regional assignment of AK-1 to 9q34. *Hum Genet* 34:35–43
45. Schmickel RD (1986) Contiguous gene syndromes: a component of recognizable syndromes. *J Pediatr* 109:231–234
46. Kimberling WJ, Fulbeck T, Dixon L, Lubs HA (1975) Localization of spherocytosis to chromosome 8 or 12 and report of a family with spherocytosis and a reciprocal translocation. *Am J Hum Genet* 27:586–594
47. Kimberling WJ, Taylor RA, Chapman RG, Lubs HA (1978) Linkage and gene localization of hereditary spherocytosis (HS). *Blood* 52:859–867
48. Bass EB, Smith SW Jr, Stevenson RE, Rosse W F (1983) Further evidence for location of the spherocytosis gene on chromosome 8. *Ann Intern Med* 99:192–193
49. Mitelman F (1998) *Catalog of chromosome aberrations in cancer*. CD-ROM, Wiley, New York
50. Mitelman F, Mertens F, Johansson B (1997) A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nat Genet Spec* 15:417–474
51. Lupski JR, Stankiewicz P (eds) (2006) *Genomic disorders: the genomic basis of disease*. Humana Press, Totowa, NJ
52. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al (2006)



- Global variation in copy number in the human genome. *Nature* 444:444–454
53. Kallioniemi A, Kallioniemi OP, Piper J, Tanner M, Stokke T, Chen L, Smith, HS, Pinkel D, Gray JW, Waldman FM (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Nat Acad Sci USA* 91:2156–2160
  54. Chung EK, Yang Y, Rennebohm RM, Lokki M-L, Higgins GC, Jones KN, Zhou B, Blanchong CA, Yu CY (2002) Genetic sophistication of human complement components *C4A* and *C4B* and *RP-C4-CYP21-TNX* (*RCCX*) modules in the major histocompatibility complex. *Am J Hum Genet* 71:823–837
  55. Le Marechal C, Masson E, Chen J-M, Morel F, Ruzsiewicz P, Levy P, Ferec C (2006) Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat Genet* 38:1372–1374
  56. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, et al (2005) The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307:1434–1440
  57. Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Robertson-Lowe C, Marshall AJ, Petretto E, et al (2006) Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* 439: 851–855
  58. Jongmans MCJ, Admiraal RJ, van der Donk KP, Vissers LE LM, Baas BF, Kapusta L, van Hagen JM, Donnai D, de Ravel TJ, Veltman JA, et al (2006) CHARGE syndrome: the phenotypic spectrum of mutations in the *CHD7* gene. *J Med Genet* 43: 306–314
  59. Lubs HA Jr (1969) A marker X chromosome. *Am J Hum Genet* 21:231–244
  60. Magenis RE, Hecht F, Lovrien EW (1970) Heritable fragile site on chromosome 16: probable localization of haptoglobin locus in man. *Science* 170:85–87
  61. Lifton RP, Gharavi AG, Geller DS (2001) Molecular mechanisms of human hypertension. *Cell* 104:545–556
  62. Wilson FH, Disse-Nicodeme S, Choate KA, Ishikawa K, Nelson-Williams C, Desitter I, Gunel M, Milford DV, Lipkin GW, Achard JM, et al (2001) Human hypertension caused by mutations in *WNK* kinases. *Science* 293:1107–1112
  63. Chang Y-PC, Liu X, Kim JDO, Ikeda MA, Layton MR, Weder AB, Cooper RS, Kardia SLR, Rao DC, Hunt SC, et al (2007) Multiple genes for essential-hypertension susceptibility on chromosome 1q. *Am J Hum Genet* 80:253–264
  64. Kajiura K, Berson EL, Dryja TP (1994) Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and *ROM1* loci. *Science* 264:1604–1608
  65. Nadeau JH (2001) Modifier genes in mice and humans. *Nat Rev Genet* 2:165–174
  66. Katsanis N, Ansley SJ, Badano JL, Eichers ER, Lewis RA, Hoskins BE, Scambler PJ, Davidson WS, Beales PL, Lupski JR (2001) Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* 293:2256–2259
  67. Schlingmann KP, Konrad M, Jeck N, Waldegger P, Reinalter SC, Holder M, Seyberth HW, Waldegger S (2004) Salt wasting and deafness resulting from mutations in two chloride channels. *N Engl J Med* 350:1314–1319
  68. Draper N, Walker EA, Bujalska IJ, Tomlinson JW, Chalder SM, Arlt W, Lavery GG, Bedendo O, Ray DW, Laing I, et al (2003) Mutations in the genes encoding 11 $\beta$ -hydroxysteroid dehydrogenase type 1 and hexose-6-phosphate dehydrogenase interact to cause cortisone reductase deficiency. *Nat Genet* 34: 434–439
  69. Saunders AM, Schmader K, Breitner JCS, Benson MD, Brown WT, Goldfarb L, Goldgaber D, Manwaring MG, Szymanski MH, McCown N, et al (1993) Apolipoprotein E epsilon-4 allele distributions in late-onset Alzheimer's disease and in other amyloid-forming diseases. *Lancet* 342:710–711
  70. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261:921–923
  71. Becker KG (2004) The common variants/multiple disease hypothesis of common complex genetic disorders. *Med Hypotheses* 62:309–317
  72. Harada K, Toyooka S, Maitra A, Maruyama R, Toyooka KO, Timmons CF, Tomlinson GE, Mastrangelo D, Hay RJ, Minna JD, et al (2002) Aberrant promoter methylation and silencing of the *RASSF1A* gene in pediatric tumors and cell lines. *Oncogene* 21:4345–4349
  73. Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301:89–92
  74. Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S, Wu Y, He X, Powe NR, Feinberg AP (2003) Loss of *IGF2* imprinting: a potential marker of colorectal cancer risk. *Science* 299:1753–1755
  75. Burchfield RW (1991) *Unlocking the English language*. Hill and Wang, New York
  76. Murray KME (1977) *Caught in the web of words: James AH Murray and the Oxford English Dictionary*. Yale University Press, New Haven, CT
  77. Buckle VJ, Edwards JH, Evans EP, Jonasson JA, Lyon MF, Peters J, Searle AG, Wedd NS (1984) Chromosome maps of man and mouse II. *Clin Genet* 26:1–11
  78. Searle AG, Peters J, Lyon MF, Edwards JH, Buckle VJ (1987) Chromosome maps of man and mouse, III. *Genomics* 1:3–18
  79. Searle AG, Peters J, Lyon MF, Hall JG, Evans EP, Edwards JH, Buckle VJ (1989) Chromosome maps of man and mouse IV. *Ann Hum Genet* 53:89–140
  80. Nicholas FW (2003) *Mendelian Inheritance in Animals (OMIA): a comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals*. *Nucleic Acids Res* 31:275–277
  81. McKusick VA (1993) *Medical genetics: a self-instruction guide and workbook based on Mendelian Inheritance in Man*. Johns Hopkins University Press, Baltimore
  82. Antonarakis SE, Beckmann JS (2006) Mendelian disorders deserve more attention. *Nat Rev Genet* 7:277–282