Minding your p's and q's

BROWSING MENDEL'S LIBRARY:
APPRECIATING THE ENORMOUS HETEROGENEITY
OF HUMAN GENOMES

Michael W Kita, MD

Daniel Dennett, in a 1995 book he provocatively called *Darwin's Dangerous Idea*, asks his readers to imagine the "library of all possible human genomes." He calls this collection the "Library of Mendel" and pictures it as containing the collection of all possible human DNA sequences— not just the sequences of all six billion people currently alive today, but every possible sequence-variation imaginable. Since the human haploid genome is three billion nucleotides long, and any position in that nucleotide chain could be held by one of four different DNA-bases (T, A, G, or C), the number of possible sequence-variations is on the order of four to the three billionth power.

Such a collection would contain sequence-variations that are actually observed (alive today or compatible with life) and ones that have never been seen (incompatible with life, or not within the meiotic potential of contemporary genomes). Of the sequences compatible with life, certain variations would be so "neutral" relative to some more prevalent or "conserved" sequence, as to be indifferent to health or longevity. And others would not.

The fact that ontogeny and phenotypic expression are the consequence of our diploid genetic endowment adds a further layer of complexity to genomic variation. Some biologic functions seem to require biallelic expression, while others are (or can sometimes be) monoallelic. In addition to the sequence differences between maternal and paternal homologues, functional differences such as imprinting, lyonization, *trans* effects, and dominance patterns also influence overall phenotypic expression of the diploid heritage.

To make matters more complicated, there are possible variations of the human genome which are longer (one nucleotide longer, two nucleotides, etc.) or shorter (by one, two, or more) than the three-billion nucleotide fixed-length chain modeled above, introducing more sequence-variability and more potential phenotypic variability into Mendel's Library. Humans have their three-billion nucleotides parsed into chromosomes, and depending on which chromosomes are affected by length-discrepancies and where insertions and deletions occur, there are all manner of aneuploidies and aneusomies, some karyotypically obvious and others quite subtle, that can be expressed in living phenotypes.

While "Mendel's Library" of human-genome sequences would indeed be enormous, and while DNA provides the basic instruction-set for fabricating a human being and maintaining it biochemically intact over a lifetime, the potential for human diversity is even greater than the genomic diversity. This is because the DNA sequences themselves are not fully self-determining, but are *intermediated* by a cascade of processes (replication and transcription fidelity, transcript-processing, translation, post-translational modification) and *circumstances* (substrate availability, degradation and turnover rates, promotion and inhibition, macro- and micro- environments) that influence moment-to-moment behavior of complex systems.

And the final "phenotypic" expression of a genome— morphologically and physiologically— is both *holistic* (integrating and

---

**Table One**

**Mendelian "qualifiers"**

Various terms have evolved to describe why a Mendelian disorder doesn't always exhibit in a Mendelian pattern, why non-Mendelian disorders can sometimes be phenotypically mistaken for Mendelian ones, and how intermediate or variable phenotypes occur.

- incomplete dominance
- codominance
- incomplete penetrance
- delayed penetrance
- variable expressivity
- sex-limited
- sex-influenced
- recessive epistasis
- uniparental disomy
- mosaicism
- imprinting

*Vice President and Senior Medical Director for UNUM Life Insurance in Portland, Maine*
summatitng the structure and function of some 10 trillion cells in a given adult) and dynamic (able to show adaptation and compensatory changes over time—responses which preserve vital homeostasis—but also subject to compounding and cumulative changes over time, in ways that may be expressed as disturbed functioning, ill-health, or premature mortality). Gene therapy may impact aspects of this holistic and dynamic balance, but how much and how soon remains to be seen.

Human molecular genetics has undergone explosive growth, and keeping pace with new developments is an enormous task.

Each new discovery in genetic research challenges medical directors and other physicians to understand what is being reported, and what its implications are. The basic questions about a new finding are generally quite simple: "What does it say?" and "What does it mean?" but the answers often are not. This is partly because the vocabulary of genetics has become more complicated, and partly because the social context in which interpretation is occurring is evolving as well.

Genetics began as a descriptive science with useful rules based on grossly observable phenotypes which permitted fairly dependable generalizations about pedigrees and populations. With the revolutionary advances of molecular biology, genetics is becoming a more predictive science based on genotypes leading to inferences about individuals and families. But its predictions are most often statements of probability or likelihood, as opposed to certainty or inevitability. And while the risk implications of genetic conditions can still be dealt with actuarially in

![Figure One](image)

**Figure One**

**Historical view: “central dogma”**

1. Genetic expression as a two-step consequence of transcription and translation.
2. Replication, transcription, and translation are unidirectional (non-reversible) processes: DNA directs its own (semi-conservative) replication and one-way transcription, and mRNA directs protein synthesis.
3. One gene: one protein.

![Figure Two](image)

**Figure Two**

**Image and mirror-image in DNA, mRNA, and protein-product**

<table>
<thead>
<tr>
<th>Correspondences</th>
<th>Complements</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' → DNA strand → 3'</td>
<td>3' → DNA strand → 5'</td>
</tr>
<tr>
<td>5' → mRNA → 3'</td>
<td>3' → cDNA → 5'</td>
</tr>
<tr>
<td>amino → protein → carboxy terminal</td>
<td>terminal</td>
</tr>
</tbody>
</table>

1. The 5' → 3' strand of human double-stranded DNA is the DNA equivalent of the transcribed mRNA and therefore is considered the "sense" strand. While it is itself the untranscribed strand, it is more directly comparable to the mRNA and therefore 5' → 3' DNA is the strand whose sequence is generally reported in the genetics literature as the coding sequence.

2. The 3' → 5' strand of genomic DNA is the so-called "anti-sense" strand. It is the actual template for mRNA and thus is the transcribed strand, but its sequence is complementary to the mRNA.

3. cDNA ("complementary DNA") is the immediate single-strand DNA complement of processed mRNA. It is the result of reverse-transcription of processed mRNA and thus is complementary to both processed mRNA and the exon-sequences of genomic DNA. This makes it suitable for a cloning vector (after first making the cDNA double-stranded again) or a DNA-probe.

4. Topologically, the 5' (upstream) ends of genomic DNA and mRNA correspond to the amino end of the translated protein, and the 3' (downstream) ends, to the carboxy end.

5. Because the sense strand, the mRNA, and the protein product correspond to each other, knowing the "code" or sequence of any one of them is informative about the sequence of the two others, and forms the basis of various gene isolation and gene-expression research techniques.
these terms, society seems ambivalent about how to deal with genetic information in an insurance context. Is health insurance a right of citizenship or a market good? Should it become a form of social insurance, or can the private insurance mechanism continue to provide the coverage in some traditional or modified way? How can or should genetic information be used in underwriting and risk classification, and for what kinds of insurance and risk decisions?

Clinical medicine must approach genetic information scientifically for diagnosis and prognosis; clinicians must use it cautiously and selectively after the appropriate weighing of factors like pre-test likelihood, sensitivity, specificity, cost and benefits: and with duly informed consent. Insurance medicine must bring the same scientific rigor to its proper uses of genetic information for actuarially-sound risk-assessment purposes and must do so with an identical respect for the individual's autonomy and confidentiality and the integrity and equity of the process.

In February of 1993, AAIM and ACLI jointly sponsored a symposium in Fairfax, Virginia on "genetic issues in insurance medicine," along with the American Society of Human Genetics and the HIAA. The symposium (published as a JIM supplement in the Summer of 1993) helped to establish both a common dialogue and a common framework for understanding genetic issues in an insurance medical context. Since 1993, genetic advances have been the subject of presentations at AAIM and ACLI scientific sessions.

As this current issue of JIM is being published, the ACLI and AAIM will jointly be sponsoring another symposium to be held in February of 1996 in Atlanta. Given the interest and the need, it is timely that a theme of this issue of JIM should be genetic issues. The genetics articles in this issue summarize progress in an insurance medical context. Since 1993, genetic advances have been the subject of presentations at AAIM and ACLI scientific sessions.

Like most gene-loci, the brcal locus begins with a promoter region (P), followed by an untranslated region (UTR) which contains the transcription initiation site (I) and leader-elements. This is followed by a series of exons and introns (the introns are not shown here) concluding with a transcription termination site (T). Typically a UTR "trailer" sequence and the poly-A tail sequence precede the T-site.

The brcal gene codes for a transcription factor that may be important in estrogen and progesterone mediated growth regulation. It is a large gene (100kb) coding for a large protein (1863 amino acids) with two zinc-finger domains characterized by molecular biologists as a "RING finger motif." Such domains are often seen in vital regulating sequences. When abnormalities of such proteins are produced, the pathophysiologic result can be transcription dysregulation and defective "tumor suppression." Some mutations in brcal have, not surprisingly, been seen in familial clusters of breast and ovarian cancer (leading to the brcal- "breast cancer 1" designation for this sequence, since some of the gene's pathologic manifestations have preceded description of the functions of the normal gene).

At present, there are over 60 mutations of brcal known, many of which are benign polymorphisms. But five critical mutations account for 35 percent of the genomic variation and are associated with a significant portion of brcal-related breast cancer. These are shown as sites A, B, C, D, and E.

**Location**  **Mutation**  **Interpretation**

- **Site A**  - 185 del AG A two pair base-deletion beginning at nucleotide 185 in exon 1
- **Site B**  - cyst0ly An amino acid substitution (glycine for cysteine) at codon 61 in exon 6
- **Site C**  - 1294 del 40 A 40 nucleotide deletion at nucleotide 1294 in exon 11
- **Site D**  - 4184 del 4 A 4 nucleotide deletion at nucleotide 4184 in exon 11
- **Site E**  - 5382 ins C A single nucleotide insertion (cytosine) at nucleotide 5382 in exon 20

Research suggests that brcal's protein product is not mutated by truncation in most breast cancer cell lines, but probably exerts its deleterious effects by mis-sense mutation (defective transcription regulation) or mis-location (aberrant sub-cellular localization, mainly to cytoplasm instead of the nucleus).
hacking, but is subject to some mutational dubbing (both germ line and somatic) and to splicing, mixing, and alteration during meiosis, mitosis, or defective repair.

Human genetics, in simple terms could be considered the science of heredity and human variation. But, “variation” includes both differences that matter biologically, and ones that do not. And, heritable traits may be fully transmitted, or only variably expressed. Human genetics may just as easily be called the science of exceptions, and this would be a good general characterization of the discipline on at least two levels: “exceptions” in the sense that mutations diverge from some expected genotype, and also in the sense of genetics seeming to have almost as many exceptions as rules when it comes to phenotypic expression. Genetics has evolved its own special language to deal with the diversity that underlies its exceptions. (Table One provides a partial list.) This is not at all surprising, given the complexity with which the human genome is organized, and the complexity of processes which underlie transmission and expression of genetic traits.

Consider that it used to be thought that “genes” were units of heredity strung like beads on chromosomes. Science now appreciates that while there are as many as 50,000 to 100,000 human genes, they are not evenly (or even randomly) distributed on the chromosomes, but tend to be localized to GC-rich areas or the telomeric ends of chromosomes. And science now reckons that only ten percent or so of the genome codes for functional genes, while the remainder codes for transfer and ribosomal RNA transcripts, or represents centromeric heterochromatin, transposed sequences, tandem arrays, pseudogenes and other interspersed elements.

Consider that it used to be thought that a gene was a continuous stretch of DNA and that this stretch was expressed as a protein, (each DNA triplet corresponding to amino acid) after transcription to a messenger RNA. Science now knows that genomic DNA has untranscribed sections representing such things as initiation and promotion sites, and the primary mRNA transcript has untranslated sections such as introns, leaders, and tails. The “cen-
tral dogma” is no longer DNA → transcription → mRNA → translation → protein, but a more intricate cascade involving post-transcriptional processing, and post-translational glycosylation and other modifications. See Figure One.

Consider that it used to be thought that one gene corresponded to one protein-product. Science now finds that certain “nested genes may code for several transcription products (by mechanisms varying from the truncation of transcription to staggered reading-frames) and that translated transcripts may yield a series of nested or cleaved protein-products (like the ACTH/MSH/endorphin series and the kinins and clotting factors), not to mention the further heterogeneity introduced by multimers (e.g. protein-protein physical interactions) and differential side chains and prosthetic groups (protein-protein chemical interactions). It is entirely possible that 100,000 genes provide the instructions for substantially more than 100,000 proteins, and that the intracellular and extra-cellular milieu that results is a dynamically rich source of phenotypic variability.

Consider that it used to be thought that humans carried 22 autosomes each, in pairs, and either two X-chromosomes (if female) or an unmatched pair of X and Y (if male), and that the autosome expressed their genes in allelic pairs, and (after lyonization) the sex-chromosomes expressed themselves hemizygously. Science now understands that X and Y in fact have pseudo-autosomal regions that pair during meiosis, and that genomic imprinting leads to allelic suppression of various genes at certain times (e.g. embryogenesis) or under certain conditions.

Consider that terms like “cistron” and “operon” have evolved to more adequately characterize functional properties of genomic genes, and cDNA (complementary DNA) to characterize the exon-unique sequence that used to be imagined as the obligatory template for the primary mRNA transcript. See Figures Two and Three.

Genetics is indeed the science of heredity and human variation, of the transmission and expression of genetic traits. When viewed broadly in this fashion, genetics encompasses heritability from parent to child (meiotic inheritance patterns) and from parent-cell to daughter-cell (mitotic inheritance). Mutational events arising during or propagated through meiotic processes include Mendelian inheritance patterns, cytogentic and karyotypic abnormalities, and (albeit less clearly or predictably) polygenic and familial-genetic patterns. Mutational events arising in nongerm cells and propagated mitotically (often by clonal proliferation and failed apoptosis) include somatic mutations such as most forms of “acquired” cancer.

Traditional classifications of genetic disorders have generally been into four broad categories: cytogenetic (aneuploidies and aneusomies), Mendelian (autosomal and X-linked, dominant, and recessive), polygenic or multifactorial disorders, and somatic mutations. See Figure Four. When viewed this way, the “hereditary” aspect of these genetic disorders can be a bit ambiguous. See Figure Five. Some cytogenetic disorders (e.g. Down’s syndrome) affect neither parent, yet are the consequence of transgenerational events (e.g. non-dysjunction or translocation) involving meiosis and abnormal gametes. Whether the disorder is further heritable to the next generation depends on the subsequent reproductive maturity and germline representations of the defect. Other cytogenic disorders (e.g. the Philadelphia chromosome in CML) are somatic mutations, transmissible mitotically to daughter cells but not transgenerationally through germ cells. Some “hereditary” conditions (e.g. autosomal recessive disorders, in which both parents are phenotypically normal, but are “carriers”) do not show a “vertical” pattern of inheritance but produce a pedigree which would look, in a single affected offspring, like a “sporadic” mutation. And if the affected offspring successfully reproduced, the next generation would appear to be unaffected (e.g. “no apparent hereditary consequence”) unless of course the mate were by chance a “carrier” for the same condition.

Some disorders may be a mixture of hereditary mutations and somatic mutation (e.g. the first (inherited) mutation affects one allele in a single-hit fashion, with the somatic mutation knocking out the other allele in a second-hit event). Other inheritance patterns may show increased familial incidence of traits due to multiallelic-polygenic expression, or complex-interactive polygenic expression.

More recent classifications of genetic disorders have, for the reasons mentioned above, divided inheritance into simple traits (Mendelian monogenic inheritance) and complex traits. See Figure Six. The latter would typically include pleiotropism (multiple phenotypes from the same genotype), incomplete penetrance, phenocopy, genetic heterogeneity (same or similar phenotypes from different genotypes), variable expressivity, polygenic inheritance, bilineality (unmarked homozygosity), mitochondrial inheritance, imprinted hemizygosity, and expansion of trinucleotide repeats.

Additional factors that confound simple inferences about genetic mechanisms in disease include the bipolarity of many forms of cellular regulation (e.g. some disorders can be due to “activation of promoters” or growth factors while a similar manifestation may result from “inhibition of suppressors”); the effects of compensatory factors or complementation; the effects of comorbidity or confounding factors; the effects of subcellular localization and compartmentalization of protein action; and the effects of timing. Ultimately, a genetic defect constitutes too much or too little of a normal (or abnormal) protein in particular places (cells, tissues) at particular times (embryogenesis, childhood, adolescence, adulthood, senescence). See Figures Seven and Eight.

Genetic heterogeneity (locus heterogeneity) describes the potential array of mutational changes within a single gene-locus. Mutations can be of several general types (see Figure 9) with corresponding wide differences in biologic effect ranging from “virtually none” (so minor or compensatable as to be essentially “silent” or “benign”) to drastically deleterious (severely morbid, or prematurely mortal).
Benign mutations are commonly called polymorphisms (alternative forms of a “healthy” or functional gene), although the term has a stricter technical meaning in population genetics of at least one percent gene-frequency (approximately two percent heterozygote prevalence). Perhaps “permuation” would accurately characterize a benign or apparently innocent gene-mutation that is not so frequent as to represent a polymorphism in the narrow sense. Deleterious-mutations can be biologically adverse in a variety of ways, from disease-producing to disease-predisposing, depending on the severity of functional disturbance precipitated by the particular protein-product abnormality in its full cellular and organismic expression.

The coding changes or coding errors caused by mutations fall into several broad categories. See Figure 10. Generally speaking, they are either length-mutations with gain or loss of genetic material, or point mutations with alteration of the genetic code, but no gain or loss of genetic material. Some mutations involve whole codons while others disrupt codons and result in frame-shifting.

Given the vast number of mitoses that occur in a lifetime— to create a 10-trillion-cell adult from a single-celled zygote (See Figure 11), and then to repair, replace, and maintain various cells—it is not surprising that the majority of mutations occur in somatic lineages. The term “somatic mutation” generally describes a mutation occurring in a somatic cell (as opposed to germ-line cell) although the term often has a further connotation of “acquired” in life, and not received from mother or father genetically.

Somatic-cell mutations thus are inheritable to daughter-cells, but since such mutations do not generally involve germ-line cells, they are not transmissible to the next generation. However, the effects of somatic mutations can still be of great consequence to the individual:

a) The earlier in life they occur.
b) The more “upstream” their location in a gene-locus, if the mutation produces frame-shift or a premature stop-codon.
c) The more synergistic the mutation is with the deleterious effects of already existing mutations (for instance, if a transgerational inherited mutation were already present in the opposite allele).
d) The more critical the affected peptide-domain is to the full and final functional expression of the protein product.
e) If the mutation affects gene-regulation, and up-regulation occurs (constitutive, as opposed to facultative, expression) then the effects may be serious.
f) Or, if the mutation produces down-regulation, (or switches to a gene “off”), and the cell or body requires the protein-product in biallelic or constitutive amounts.
g) Or, if the mutation affects a critical editing or repair enzyme.

The DNA sitting in a cell’s nucleus is not completely protected from harm by its compartmentalized location, its super-coiling, its double-stranded state, or its histone/nucleoprotein jacketing. Rather, it is subjected to chemical mutagens from diet and environment (e.g. base-analouges, alkylating agents, intercalating agents, deaminating agents, etc.) and exposed to physical agents (e.g. electromagnetic radiation—especially in the X-ray and UV ranges—and thermal factors). Not all locations in the genome are equally likely to undergo mutation, and evidence suggests that “transition” mutations (C→T and G→A) are among the more frequent, and that the “CG” pair is a genomic “hotspot” with regard to mutational potential. In the absence of effective repair/editing enzymes, it is estimated that in each human cell spontaneous deamination would degrade 100 or more genomic cytosine sites (C→U and 5-methyl-C→T) and 5,000 or more purine sites would undergo “transition” or “transversion” mutations each day.

Even in the absence of these various mutagenic influences, the genome would be at risk of “spontaneous” mutation via replication error. Although replication proceeds extremely fast (20 base pairs per second), the introduction of an incorrect nucleotide is a rare event (one in 10 million base pairs) made rarer still by “editing” enzymes that correct many replication-faults. Overall, replication errors occur only once in 10 billion base pairs per cell division (about once in every other completed mitosis). However, since there are perhaps 10^15 cell divisions during a human lifetime, replication errors result in thousands of new mutations at virtually every nucleotide position in the genome, in some cell within the body.

Mutational events also affect germ-line cells. Germ-line mutations can arise from replication errors involving primordial germ-line cells, from mutagenic influences on germ-line cells or directly on gametes, and from flawed meiosis. In the female, Oogenesis is completed in fetal life, prior to birth, and consists of approximately 22 mitotic divisions and one suspended meiotic division. (Ova in the female do not complete meiosis II until fertilization occurs). Spermatogenesis also involves numerous replications:

1) about 30 mitotic divisions between embryogenesis of primordial germ cells and puberty,
2) about 20-25 division cycles per year from puberty though adult life, and
3) a final replication cycle creating 4n bivalents just prior to meiosis I.

Completion of spermatogenesis then involves a final two-stage meiotic division. Mutations involving spermatozoa and ova can arise from mitotic (germ-line-somatic) or meiotic (germ-line-gametic) errors. The enormous number of sperm cells produced per ejaculation is such that it has been estimated that one in 10 sperm may carry a new deleterious mutation.

However, given that a germline mutation requires both transmission (to progeny) and expression (survival of those progeny) to be of heritable significance, germline mutation rates per generation are far lower, on the order of 10^-9 to 10^-8 per locus. Still, given that there are approximately 100,000 genes in the
human genome, this means at least one person in 10 is likely to have received a mutated gene from one or the other parent.

When mutations do occur, how are they detected, how are they localized to specific chromosomes and genes, and how is the function of the gene (and dysfunction of the altered gene) deduced? See Figure 12. Because of the "correspondences" and "complementarities" afforded by the central dogma (Figures One and Two), one could start anywhere in the circle of relationships connecting genotype to phenotype and work forward or backward.

If working from the phenotypic end of the spectrum (see Figure 13), it is advantageous to have a precise phenotype. For example, CAD or hypercholesterolemia are insufficiently precise descriptions of phenotype from which to work backwards to a genotype, since they would include potentially several single-gene disorders affecting cholesterol metabolism, plus phenocopies, plus polygenic and multifactorial disorders. But, "premature familial CAD due to an LDL anomaly with hypercholesterolemia" begins to narrow the phenotype down to one (familial hypercholesterolemia) which has an informative pedigree (autosomal dominant), and a functional abnormality of specific protein (the LDL receptor).

Such a protein, or its key functional domains, can be sequenced, and possible mRNA and DNA coding-equivalents inferred. Probes can then be devised and the gene localized. This bottom-up approach and variations on this theme— which presumely a spe-

| Figure Nine |
| Transcription: mutational consequences |
| Gene | mRNA | functional protein | healthy phenotype |
| Polymorphic mutation | altered mRNA | functional polymorphic protein | indifferent to overall health status |
| Deleterious mutation | altered mRNA or no mRNA | non-functional | disease (with single-gene disorder behavior) |

| Figure 10 |
| Mutational mechanisms:† |

**No frameshift— whole-codon changes**

1) Point mutations: transition (A→G, T→C) and transversion (purine→pyrimidine)
   a) (intra-exonic) amino acid substitution: some are silent/neutral, others affect active site (Km or Vmax), conformation (allostery), membrane attachment, inducer/inhibitor binding sites, and result in "mis-sense" of various kinds
   b) (intra-exonic) premature stop codon: truncated protein, producing sense, mis-sense, or nonsense, depending on how upstream it occurs
   c) (non-exonic) splice site mutations, etc: variable effects, mis-sense to nonsense

2) Length mutations: whole codon (triplet) or multiple-codon (a number of nucleotides divisible by three), insertions or deletions, (including trinucleotide repeat insertions) with preserved reading frame
   a) (intra-exonic) insertions or deletions: variable effects depending on location and extent of alteration of final protein-product
   b) (non-exonic) insertions or deletions in introns, leaders, or tails: variable effects, depending on type of derangement of various recognition sites

**Corrected frameshift**

Insertion or deletion of three nucleotides (or multiple of three) at a location other than between two codons: e.g. if a length mutation results in the deletion of the third nucleotide in one codon, and the first two nucleotides in the next codon (as in the D508 mutation in cystic fibrosis), two adjacent reading frames are affected, but the net result is no propagated frameshift. Rather, there is a substitution of one amino-acid for the original two, and a shortening of overall peptide length by one amino acid. (The biologic effect of which will depend on the criticality of the affected locus to protein-function).

**Frameshift— fractional-codon/variable-length changes**

1) Single nucleotide and two nucleotide insertions and deletions: affects everything downstream from site of occurrence, with re-write of amino acid sequence (with possible creation of an interpolated stop codon).

2) Non-triplet (nucleotides not a multiple of three) insertions or deletions include two, four and five nucleotide VNTRs except if multiple of three): sequence re-write, with mis-sense or nonsense depending on how upstream the genomic insertion occurs.

**Partial or whole-gene duplications or deletions**

Unequal crossing over, chromosomal translocations, chromosome breakage, aneuploidies: variable effects, depending on location and extent of gene-loss or redundancy.
to incomplete penetrance and subject the offspring? which may be a Mendelian trait but is also pedigreed. If the father goes on to manifest the later condition (blackened square) lifetime of some family zygote (Z) adding offspring to the length, contribute meiotically-produced gametes (G) to a new cally unfolding process. Two adults, after a mitotic V a dynami- ~"-" A pedigree is a snapshot-in-time of Genetics: The Science Of Heritable Differences

Contig: contiguous stretches of cloned DNA (linked or overlapping) or a DNA fragment or a karyotypic abnormality) that occurs informatively in family members, then one can close in on the gene through linkage analysis, chromosome walking, knock out studies, and the like. When the likely locus is discerned, DNA from that location can be cloned, and sequenced. The sequenced gene can then be expressed and its functional implications deduced at a molecular level via its ultimate protein-product. Approaches such as this are termed “positional cloning.”

Approaches such as these allow for mapping of the human genome. They yield information about the physical map and about the genetic map. See Figure 14. The physical map of the genome gives distances as measured in base pairs (nucleotides) and at its greatest level of detail, names the individual nucleotides (specifies exact sequences).

The genetic map measures distances in terms of recombination frequencies between linked markers. A typical unit of measurement for a genetic map is centimorgans (cM) where one centimorgan equals a recombination frequency of one percent (recombination occurring in one meiosis out of every 100). The genetic map and the physical map have significant correspondences, much as a street map and a subway map of a city relate to each other. Roughly speaking, one centimorgan on the genetic map corresponds to about one million bases on the physical map (one cM = one Mb). However, the two maps are not identical. Recombination frequencies vary in different regions of the chromosomes. Consequently, while the human genome is estimated to be 3,000 Mb in length, it will not exactly equal 3,000 cM in length.

Messenger RNA is the connecting link between genotype (DNA) and phenotype (protein-expression). At whatever point in the investigation of genes and proteins quantities of mRNA become extractable or synthesizable, “functional cloning—” the use of mRNA and its direct counterpart cDNA to study functional gene expression— become possible. Figure 12 shows some of these relationships.

A gene is DNA. Logically then, a genetic test is a test on DNA or a test which uses DNA to tell something. But because of the interconnection between DNA, mRNA and protein-product, and the larger but often blurrier relationships between genotypes and phenotypes (See Figures 13 and 15.), establishing a simple or universally agreed-upon definition of what is and what isn’t a genetic test is not easy. Is a genetic test any test used to assess a suspected genetic condition, or is it only a tests which involve assaying DNA in some fashion? See Figure 16. If a “genetic test” is only one which somehow defines “genotype,” where does geno-
type end and phenotype begin? See Figure 13. Is "genotype" the domain between DNA and mRNA (the "code" and the "transcript") and phenotype the domain between protein-product and whole-person functional manifestation? Is "genotype" information about sequence (from DNA sequence to amino acid sequence) and "phenotype" then information about function, both cellular and whole-person? If so, are genotypic tests "genetic" and phenotypic tests, not? When one is dealing with a continuum between genotype and phenotype, the distinction between the two is apt to be arbitrary, and may be influenced by societal attitudes toward the subject, as much by scientific categorizations.

The interpretation of the medical information in genetic tests—however defined—needs to be accurate. It needs to be understood in all its conditional and relativistic fullness. It needs to be handled with due respect for principles like autonomy and confidentiality. And the information needs to be scientifically applied to appropriate risk-decisions in a fair and equitable manner. When this is the case, there can be harmony between private insurance as a risk relief mechanism and societal values regarding the proper way to treat individuals and groups. If the necessary trust is there, or if dialogue establishes the necessary trust, debate over the proper uses of genetic information can only affirm the legitimate role of private insurance as a social institution. This relationship has weathered past storms around other types of sensitive personal information including psychiatric history, family history, substance abuse information, and HIV test results.

The more one studies in Mendel’s library, the more appreciation one has for the heterogeneity of human genomes. Rather than one master-sequence, there are likely to be many benign permutations and polymorphisms. Even when potentially deleterious, many mutations will express themselves variably, according to a number of interacting and modulating factors. And even when a serious mutation constitutes a significant risk-factor, disease-susceptibility, or disease-predisposition, its actuarial significance will vary with age and competing risks. One might say that it causes a mutation in μ: the force of mortality that shapes the Gompertz curves of different risk-groups.

### Bibliography


---

### Figure 13

**The Spectrum From Genotype to Phenotype**

<table>
<thead>
<tr>
<th>Genotype: DNA sequence</th>
<th>Huntington's Disease</th>
<th>Familial Hypercholesterolemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene sequence</td>
<td>DNA (sequence)</td>
<td>DNA (sequence)</td>
</tr>
<tr>
<td>gene marker</td>
<td>Gene probe</td>
<td>(7)</td>
</tr>
<tr>
<td>transcript</td>
<td>mRNA</td>
<td>(mRNA)</td>
</tr>
<tr>
<td>translational product</td>
<td>Huntington</td>
<td>LBD receptor</td>
</tr>
<tr>
<td>down-stream biochemical</td>
<td>brain dysfunction</td>
<td>LDL</td>
</tr>
<tr>
<td>or physiologic effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>temporal effects</td>
<td>dementia &amp; motor</td>
<td>premature MI</td>
</tr>
<tr>
<td>end-stage disease</td>
<td>severe neurological</td>
<td>heart disease</td>
</tr>
</tbody>
</table>

**Phenotype: "outward" manifestation**

---

### Figure 14

**Genetic distance**

1) Pedigree distance.

<table>
<thead>
<tr>
<th>Degree</th>
<th>Relationship</th>
<th>Portion of genes in common</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>parent/child and sib/sib</td>
<td>1/2</td>
</tr>
<tr>
<td>Second</td>
<td>grandparent/grandchild and uncle/niece</td>
<td>1/4</td>
</tr>
<tr>
<td>Third</td>
<td>great grandparent/great grandchild and</td>
<td>1/8</td>
</tr>
<tr>
<td></td>
<td>first cousins</td>
<td></td>
</tr>
</tbody>
</table>

Because of consanguinity, even fifth cousins have 1/2048 genes in common, and offspring of a fifth cousin marriage will be "homozygous by descent" at approximately 24 loci. It is also interesting to note that a child never receives a parental chromosome that is exactly like either copy the parent carries. Rather, with approximately one to three recombinations per chromosome, a chromosome inherited from the mother will contain alternating stretches of the maternal grandmother’s chromosome and the maternal grandfather’s chromosome in a meiotic patchwork.

2) Map distance

<table>
<thead>
<tr>
<th>physical map</th>
<th>genetic map</th>
</tr>
</thead>
<tbody>
<tr>
<td># nucleotides</td>
<td>recombination frequency</td>
</tr>
<tr>
<td>gene-sequence</td>
<td>lod scores</td>
</tr>
<tr>
<td>kb or Mb</td>
<td>cm</td>
</tr>
</tbody>
</table>

(General relationship is 1Mb=1cM; this is only approximate)

In December 1995, a physical map of the entire human genome was constructed for the first time using sequence-tagged-sites spaced about 100kb apart. It is the highest resolution map yet created, and sets the stage for large-scale positional cloning and gene sequencing.
Figure 15
Genes Or Environment?

Diseases occur along a continuum from the "purely" genetic to the "purely" environmental.

Genes: Genetic Diseases
- Mendelian:
  - Huntington's (AD)
  - Cystic Fibrosis (AR)
  - Hemophilia (XL)
- Cytogenetic:
  - Down's

Environment: Multifactorial
- CAD (polygenic)
- Cancer (somatic mutations)
- Depression (multifactorial)
- Environmental Trauma (car accident)
- Infections (HIV)

Exceptions, e.g.,
- PKU on a diet free of phenylalanine (an autosomal recessive inherited condition whose severest manifestations are prevented by environmental dietary modification)
- CAD due to familial hypercholesterolemia (autosomal dominant inherited disease with phenotype very similar to multifactorial CAD)
- MVA due to a ruptured aneurysm in Marfan's syndrome (The car accident looks like "external" trauma, but the "cause" of the MVA is natural, not accidental).

Figure 16
When Is A Test A Genetic Test?

- What makes a test a genetic test?
- Is it the disease for which the test is being performed, or is it in the nature of the test being used?

One way to look at it
Tests for Genetic Disease
- Non-genetic Diseases
  - DNA
  - non-DNA
  - human
  - non-human
  - Direct
  - Indirect

Alternative way to look at it
Tests with DNA
- non-DNA

McKusick, VA, "Medical Genetics: Forty Year Perspective on the Evolution of a Medical Specialty from a Basic Science," *JAMA* 270:2351-2356, 1993