

# A Physician's Primer *of* Clinical Genomics

# Table *of* Contents

Preface

1. Introduction to Clinical Genomics

2. Piecing the Genetics Puzzle Together:

- *The Mystery of Life*
- *How Genes Work*
- *A Brief Timeline of the Discoveries and Concepts of Heredity and Molecular Genetics*
- *Diagnostic Genetic Methodologies*

3. Mendelian Inheritance

4. Predictive Genomics

- *Polymorphic Variation and Single Nucleotide Polymorphisms*
- *Nature vs. Nurture*

5. Case Study Examples

6. A Few Bioethical Considerations

7. Resources

Footnotes

## **Preface**

As practitioners, most of us had limited training in genetics, and a lot has changed in recent years. The field of clinical genomics now promises to revolutionize the way we practice not only primary care medicine but also every individual sub-specialty of medical practice. Being able to read the genetic code of individuals will prove to be every bit as great an advance in clinical diagnostics as the pioneering work of Robert Hooke and Anton van Leeuwenhoek with the microscope.

Although many of the advances in the past five years have been methodological, rather than clinical, these new methods of investigating the genome are now being applied to common clinical conditions, revealing novel therapeutic approaches with the promise of highly favorable outcomes. The new information is coming fast and furious, and will continue to do so for years to come, but the time to begin educating ourselves is now.

This guide places the new genomic information into a clinical context for practitioners and provides background information and defines key concepts so that the health care provider can understand the practical clinical applications of current genetic research.

# 1. Introduction to Clinical Genomics

The concept of "biochemical individuality" was first proposed by Roger Williams in 1956 to explain variability in disease susceptibility, nutrient needs, and drug responseveness among otherwise seemingly healthy people. It is only in the wake of the ongoing genomic revolution, however, that predictive genetic testing has become available to allow us to assess true biochemical individuality. For the first time, physicians can predict with increasing precision who is more likely to develop specific diseases, who will respond favorably to a particular drug or supplement therapy, or react adversely, and finally, which nutrients are optimal for a particular individual's treatment, health, and well-being.

Genetics is the scientific study of heredity, one gene at a time. Genomics is the study of genomes, or the totality of the DNA of a single species. Genomics attempts to look at the totality of all our genes as a dynamic system, interacting with and influencing our biochemical pathways and physiology. The Human Genome Project is the mapping and sequencing of the entire human genome. The first draft of the entire human genome was published in April 2001, almost exactly one hundred years after the rediscovery of Mendel's "Laws of Heredity." The human genome consists of slightly more than 3,000,000,000 nucleotides (give or take a few hundred million) and it codes for every protein and every enzyme made by the human body. Some 40,000 or so genes are thought to exist in the human genome, yet we know the function of less than half of those genes.

As primary care practitioners, we stand at a critical crossroads where increases in availability of DNA-based testing and demand by patients for genetic information and advice necessitate our need to become genetically literate. This primer will provide broad-based and novel instruction in the advances of genomics with emphasis on clinical implementation of genomic information.

## *Clinical Genomics*

Knowledge of the human genome will revolutionize the practice of medicine. Currently, three broad areas of clinically relevant genomics are rapidly advancing:

1. Mendelian inheritance
2. Pharmacogenetics
3. Predictive genomic testing for chronic diseases with multifactorial etiologies

## *Mendelian Inheritance*

Mendelian genetics, or the inheritance of traits or characters in pedigree patterns, encompasses the history of genetics prior to genomic analysis. Animal husbandry and plant hybridization are early, pre-scientific examples of using Mendelian inheritance, albeit without an awareness of the scientific principles involved, for specific purposes, such as improving wheat yield, breeding sheep with denser wool, hybridizing tomato plants with bigger fruits, domesticating dogs, and so on.

In humans, once patterns of inheritance became clearly recognizable through family pedigrees, doctors were able to offer limited genetic counseling for conditions like hemophilia, muscular dystrophies, cystic fibrosis, Huntington's disease, Down's syndrome, and the like. However, all too often, little or no treatment was available for those affected.

Advances in genetic testing have facilitated the identification of point mutations and chromosomal rearrangements, making the diagnosis of inherited disorders faster, easier, cheaper, and more accurate. It is now relatively simple to determine who will develop a Mendelian genetic disorder, who is a carrier, and who is unaffected within a family. This has greatly enhanced the effectiveness of genetic counseling. Further, pre-implantation genetic diagnosis is also rapidly becoming available as an alternative to intra-uterine genetic diagnosis (e.g., amniocentesis) to rule out single gene defects or chromosomal rearrangements. With such technology, we are able to prevent the need for termination of pregnancy by screening for genetic aberrations before implantation takes place. These and other strategies for prevention, early diagnosis, and treatment are becoming increasingly commonplace as primary care practitioners interface with specialized genetic counselors.

## ***Pharmacogenetics***

Pharmacogenetics, simply put, is the field of pharmacology that uses genomic information to find the *right* drug for the *right* person at the *right* time. All drugs in use today act on fewer than 500 known molecular targets. If only 10% of the genes in the genome represent molecular targets for drug therapy – a very conservative estimate – then the possibility exists to develop some 3-4,000 new molecular entities to combat disease – that's 8 times more drugs than are on the market today!<sup>1</sup>

Pharmacogenetics begins with functional genomics: determining the functions of genes in order to find those that make good targets for drug discovery. In most cases, genes code for proteins, and the proteins exert a physiological effect and thus are excellent candidates for development as drugs. Indeed, the first wave of genomic drugs to come on the market will belong to the conventional classes of protein, antibody, and small molecule drugs that are in wide use today. Yet the new drugs are likely to exhibit increased specificity and efficacy. Epogen™, an analogue of the natural protein that stimulates red blood cell manufacture in the marrow, and monoclonal antibodies are examples of these new types of "smart drugs" emerging on today's pharmaceutical markets.

Pharmacogenetics will also focus on tailoring drug treatment to specific individuals. Why do some people have adverse drug reactions or some show no response, while others receive substantial therapeutic benefit – all from the same drug at the same dosage? The answers are increasingly believed to lie in genetic variability from one person to the next.

How we metabolize specific drugs has a lot to do with our genetic capacity for detoxifying and eliminating drugs from the body. If a particular enzyme is missing or is overly abundant, a given drug may not be ideal for that patient. It is hoped that with thorough genetic analysis many adverse drug reactions can be avoided altogether.

Moreover, this same general principle of impaired detoxification enzymes may underlie the pathophysiology of numerous troubling modern chronic diseases like chronic fatigue

syndrome or fibromyalgia or autism. Some individuals may be genetically ill equipped to handle the added toxic burden of an increasingly industrialized world.

The ultimate endgame of pharmacogenetics, however, is likely to be therapies such as anti-sense RNA for shutting off specific gene expression and gene therapy for replacing damaged genes. The development of these therapies is only in its infancy with mixed clinical results and will probably not be clinically viable for some time to come.

## ***Predictive Genomic Testing***

Probably the largest potential area of medical intervention using clinical genomics may be categorized under the rubric of predictive genomic testing. In nature, changing just one nucleotide out of the 3 billion in the human genome allows for physiologic variation, for natural selection, and for evolution. Yet such changes are frequently disadvantageous for the health of the individual; only rarely is such a change physiologically beneficial. These single nucleotide polymorphisms (SNPs – pronounced "snips") are also believed to account for the vast amount of variation between individuals. Humans share in excess of >99.9% homology; it is the <0.1% of our genes that make us individuals.

Tremendous efforts are underway to identify polymorphisms, as they may hold the ultimate keys to truly individualized medicine. The future impact of SNP research on primary care medicine cannot be overemphasized: most fundamental physiological processes like detoxification, immune surveillance, hormone signaling, and metabolic pathways are all dependent in large part on individual genetic variation.

SNPs are also known to play a significant role in the development of many chronic diseases. By examining conditions like heart disease, neurological degeneration, and osteoporosis, we can demonstrate how genetic testing for SNPs can play an enormous adjunctive role in treating these common clinical conditions. While SNPs have been shown to increase risk of developing chronic diseases dramatically, there are also intervention strategies involving diet, lifestyle, and specific nutrient and drug intervention that are available to minimize or even eliminate the effects of the increased genetic risk. As SNP testing becomes widely available, the public demand for accurate risk assessment and therapeutic prevention strategies will grow exponentially.

One potential problem is the number of SNPs in the genome. To date, roughly 100,000 SNPs have been identified, and literally millions exist in the DNA, but the vast majority of these have no impact whatsoever on human physiology. Finding the polymorphisms that make a real difference in our physiology is the first task of predictive genomics.

To be of clinical utility, the polymorphisms identified in predictive genomic testing must meet four criteria: they must be,

- 1 **Relevant** – the only polymorphisms in the genome of interest are those that exert a significant effect on our biochemistry and physiology
- 2 **Prevalent** – given our current knowledge of the human genome, only polymorphisms that exist in a significant percentage of the population are likely to be identified in a cost-effective manner

- 3 **Modifiable** – only polymorphisms whose effects are modifiable via clinical intervention (diet, lifestyle, supplements, pharmaceuticals, and toxin exposure reduction) are clinically useful
- 4 **Measurable** – our genes do not change but our functional physiology and metabolic reserve do change. The progress of our clinical interventions for risk reduction and functional improvement must be measurable. Functional laboratory testing is the primary vehicle by which these changes may be measured.

Over 5% of colon, breast, ovary and prostate cancers are estimated to be due exclusively to mutant genes. Carriers can have a 60-80% chance of developing the disease.<sup>2</sup> But by the same token, 20-40% of carriers of particular mutations will NOT develop disease. Why? Because, in large part, diet, nutrition, and lifestyle factors can exert an enormous influence on how, or even if, a gene will express itself. The common misconception of genetic testing is that it foretells our fate. While that may be true of certain Mendelian traits, it certainly is not true of most predictive genomic testing. Rather, knowing about increased risk is the first step towards a committed and effective prevention strategy.

Similarly, SNPs in important biochemical pathways can alter the body's detoxification capacity and its ability to maintain proper immune surveillance. For example, multiple variations of cytochrome p-450 enzymes as well as glutathione-transferase and N-acetyl transferase have been identified and are known to play important roles in adverse drug reactions, drug resistance, as well as the development of complex syndromes like multiple chemical sensitivity and cancer, perhaps through an increased level of oxidative stress. Alterations in immune parameters can be identified through SNPs that affect the production of various cytokines like IL-1 and TNF- $\alpha$ . Genetic up-regulation of the production of these cytokines can lead to a TH-2 dominant state with increased incidence and severity of chronic inflammatory disorders.

The phenotypic expression of SNPs can frequently be modified through dietary and lifestyle choices, clinical nutrition, and judicious pharmacological intervention. Or at the very least alternative biochemical pathways can be supported to minimize the phenotypic effect of defective enzyme systems. Furthermore, laboratory testing is available to monitor modifications in physiologic capacity and function brought about through such interventions. This is what is commonly referred to as functional medicine or functional laboratory testing, which allows us the means to monitor the effectiveness of our treatment protocols, since, it should go without saying, a person's genetic propensities will never change. All we can do is affect the expression of those genes in the individual's biochemistry and physiology.

The specter of genetic determinism looms large in the public consciousness –most people are convinced that our genes are our fate. Nothing could be further from the truth. In fact, phenotypic expression of genomic determinants is largely modifiable. It is becoming increasingly evident that who we are as individuals is a function of both our genetic make up and the environment we subject our genes and our bodies to.

The goal of predictive genomic testing is to reveal underlying genetic susceptibility to a wide variety of clinical conditions and diseases. Every person wanting to take a proactive role in his or her health would benefit from an understanding of his or her genetic susceptibility and risk. For a few groups of patients, however, the benefits of predictive genomic testing may be especially great:

- Patients with a family history of chronic disease like heart disease, osteoporosis, cancer, allergy, or chronic inflammation, or patients who may have been adopted
- Patients with chronic conditions that have been refractory to standard treatment
- Patients with a history of prolonged toxic exposure

Health and disease lie at the intersection of our genes and our environment. Until now all we have been able to measure is pathology, function, and environment. Now, with the advent of predictive genomic testing we can measure genetic predisposition to many illnesses as well. For the first time in the history of medicine, we can begin to measure the true state of an *individual's* health. Medicine can assuredly never be the same. As a health care practitioner committed to using the latest clinical diagnostics for your patients, the time to develop a working knowledge of genomic testing and responsive intervention strategies is now. Otherwise, the future of medicine will pass by both you and your patients.

## 2. Piecing the Genetics Puzzle Together

### The Mystery of Life

Variety is not only the spice of life it is also the driving force of natural selection and of evolution. Scientists now believe that all life on this planet evolved from a single common ancestor. Why? Because all life "speaks" the same language of heredity: we call that language DNA. The same four nucleotides (or "letters," if you will) are used to write the DNA code in every creature that is now alive or has ever lived on Earth. Variation in the genetic code is the primary means of biological change and it is biological change that has meant fruitful abundance. Using the common language of DNA, Nature was able to write the code that has allowed this abundance of creatures to evolve. In this sense at least, all life on this planet is one.

The 20<sup>th</sup> century may well be remembered in medicine as the century of Genetics. In 1900, the scientific community knew literally nothing about the mechanisms of heredity, yet, one hundred years later, we had a nearly complete map of all 3 billion nucleotides that make up the human genome and a dozen or so other species' genomes mapped as well. Some fifteen Nobel Prizes in medicine and physiology were awarded to researchers in genetics, far more than to any other field of medical inquiry. And with good reason, for here, we are at the very gates of life itself, reading the epic saga of how life came to be on this planet and how it continues to be passed from generation to generation in an unending chain to infinity. Life, as difficult as it may be to define, appears minimally to involve two basic phenomena: the ability to replicate and the ability to create order. In Erwin Schroedinger's pregnant phrase, living creatures "drink orderliness" from their environment. In both cases, the key to order and to replication is information. That information is held within and passed on through our DNA.

Matt Ridley, in his riveting book, *Genome: an Autobiography of a Species in 23 Chapters*, draws a striking analogy for the human genome:

"Imagine the genome is a book.

There are twenty-three chapters, called CHROMOSOMES.

Each chapter contains several thousand stories, called GENES.

Each story is made up of paragraphs, called EXONS, which are interrupted by advertisements called INTRONS.

Each paragraph is made up of words, called CODONS.

Each word is written in letters called BASES.

There are 1 billion words in the book which makes it longer than... 800 Bibles."<sup>3</sup>

Claude Shannon, in the early 1940s, had the idea that information and entropy are opposing forces, each having an intimate connection with energy. In fact the more information a system contains, the less entropy it has, and vice-versa. Living systems defy entropy only in as much as they possess a high degree of information. That information comes from their DNA. It is the information found in DNA that allows living systems to "drink orderliness" from the rest of the universe and to produce offspring.<sup>4</sup>

Of course, the idea that "information" is the key to living beings is not particularly new. Aristotle believed that inherent in an egg was the "idea" of a chicken; so too, within an acorn was the "plan" for a future oak tree. Aristotle's ideas fostered the notion that there was a "vital force" integral to living beings. There is sweet seduction in the idea that the information implicit in living systems is a manifestation of some greater force that animates all life. That vital force is considered to be but one aspect of the Law. It is Law that informs, regulates and even intends the interrelationships of our physical reality. The very orderliness of the universe; our ability to express the world of matter in mathematical equations is evidence that there truly are Laws of Nature.<sup>5</sup> In such metaphors like "vital force" and "Law," we hear echoes of the opening passage of gospel of John, "In the beginning was the Word and the Word was with God and the word was God. The same was in the beginning with God... And the Word was made flesh...." As difficult as it may be to prove scientifically, many cannot help but believe that there are qualitative and not merely quantitative differences between living creatures and inanimate objects. Or in simpler terms, life is, and perhaps always will be, a mystery.

Yet the Word, the Law, the primordial information that intended life, cannot be DNA. It takes proteins to make DNA, and it takes DNA to encode proteins. Thus we find ourselves stumped with another chicken-and-egg conundrum. Neither DNA nor proteins can exist without the other. Protein is phenotype: metabolism, biochemistry, body, the chicken. DNA is genotype: self-replication, information, code, the egg.

Before proteins, before DNA, there was the chemical substance that links the two together even to this day. That link, the primordial source of information was, in all likelihood, RNA, ribonucleic acid. RNA, unlike DNA and unlike proteins, can replicate itself. Looking at cellular physiology, it is an RNA-dependent enzyme that takes the RNA message from the DNA to a RNA-containing protein complex (ribosomes) that then translates the message from the DNA into a protein using RNA-transported amino acids. Thomas Cech and Sidney Altman postulate that the first gene, the "ur-gene," was a combined "replicator-catalyst": it consumed chemicals around it in order to replicate itself. Not exactly life as we know it, yet these ribo-organisms could create order and replicate themselves.

Still, these ribo-organisms had an inherent problem: in an adverse environment, or even if they got too large, they would fall apart. Over some stretch of evolutionary time, it is thought that these ribo-organisms were able to translate their information to a more stable molecule, i.e., DNA, and from thence to a more stable substance in proteins. RNA maintained its existence as the critical go-between for DNA and proteins, indispensable to the replication process.<sup>6</sup> Thus, life developed into its more stable forms: DNA is transcribed into RNA that is

translated into proteins that, in turn, synthesizes and regulates DNA. And the cycle of life as we know it, oscillating between genotype and phenotype via the messenger RNA was born. So pivotal is this relationship to our understanding of cellular life that it is referred to as "The Central Dogma" of molecular biology.

## How Genes Work

DNA has three known functions in living organisms:

1. DNA replicates itself
2. DNA codes for RNA which in turn codes for proteins, the primary building blocks of the cell, the tissues, and the body
3. DNA regulates gene expression, allowing for
  - a. Cell growth
  - b. Cell differentiation
  - c. Cell replication
  - d. Programmed cell death

The structure of DNA is complementary. It is built from deoxyribose (a sugar), phosphate groups, and four nucleotides or bases: adenine, cytosine, guanine, and thymine (mercifully abbreviated to A, C, G, and T). Adenine can only bind with thymine and cytosine can only bind with guanine, producing the complementary structure. The 3-dimensional structure of DNA is like a ladder that has been twisted around its vertical axis: the deoxyribose and phosphate form the "rails" of the ladder, while pairs of A & T and C & G form the "rungs." The advantage of the complementary structure is simply that the DNA "ladder" can split with each half binding to complementary nucleotides in order to make two perfect copies of the original DNA.

This is no small project. If all the DNA in a single human cell were unraveled and stretched out into a straight line, it would measure about 2 meters (6 feet). Given the 100 trillion or so cells in your body, if all the DNA in all your cells were stretched out in a straight line, it would reach to the sun and back... a thousand times. That's rather a lot of information, even if, as we shall see, over 97% of it is junk.

The complementary binding of nucleotides to one another also allows DNA to code for RNA faithfully. RNA is structurally similar to DNA except,

1. RNA is single stranded,
2. RNA uses the nucleotide uracil (U) in the place of thymine, and
3. RNA's 3-nucleotide codons (think of them as "3-letter words") code directly for specific amino acids, allowing for the synthesis of proteins in ribosomes.

Only one strand of our DNA codes for RNA. This strand is called the "sense strand" while the other unused strand is referred to as the "anti-sense strand." (As a side note, one area of pharmacogenetics research involves the creation of anti-sense drugs that could bind to sections of DNA, preventing their transcription – effectively silencing a "bad" gene.)

Heredity is dependent on the genes found within the entire genome. Genes are those sections of the DNA that code for RNA (and subsequently protein synthesis). Only about 3% of the human genome is actually used by and for human physiology. The average gene is

about 3,000 nucleotides long, but this can vary considerably – the gene for dystrophin, the longest known in the human body, is an enormous 2 million base pairs long. The final messenger RNA made from such genes is much shorter, however.

Genes are composed of exons, portions that actually code for proteins, and long introns interspersed between the exons, that do not code for anything. Introns have been likened to advertisements in a magazine: contributing nothing to the actual storyline and constantly interrupting it. And, like some magazines, the "advertisements" found within a gene are far longer than the actual "story" itself. Indeed, most genes have far more introns than exons in them. The gene is transcribed into RNA containing a complementary copy of the entire gene but before leaving the nucleus, the RNA is shortened. The introns are excised and the exons are joined together to produce the final messenger RNA.

Messenger RNA is composed of triplet codons, or 3-letter "words," that code for specific amino acids. Mathematically, since there are 4 nucleotides in RNA, there are  $4^3$ , or 64, possible three-letter words. 61 triplet codons are specific for one of the 20 amino acids used to make proteins. The remaining 3 codons are stop codons, telling the ribosome to stop translation and release the protein. Obviously more than one codon can code for the same amino acid, but each codon is specific for only one amino acid. This allows for a precise translation of the genetic information in DNA into proteins. After release, the protein may still be modified posttranslationally: e.g., a glycoside or a lipid group may be added or the protein may be folded into its final tertiary shape to finish the specific protein being manufactured.

## ***Junk DNA***

A close look at the human genome is rather surprising. You might think that all 3 billion nucleotides serve some clearly useful purpose for our survival and health. Surely we humans are such complex creatures that 3 billion nucleotides would be necessary to explain our individuality. Nothing could be further from the truth. In fact, 97% or so of the human genome doesn't consist of genes at all: it's genetic gibberish and is known as junk DNA since it codes for nothing and has no known function. A significant minority of the junk DNA codes for complete viral genomes, although the viruses themselves are never expressed phenotypically as actual viruses. There are several thousand complete viral genomes within our human genome. Known as human endogenous retroviruses, or Hervs, these viral genomes constitute about 1.3% of the entire genome, a staggering figure when you consider human genes only constitute roughly 3% of the genome.

Most of the remaining junk DNA is classified as minisatellite DNA, characterized by the presence of short arrays of tandem repeat units – base pairs that repeat a short to medium sequence over and over and over. Minisatellite DNA is remnant viral DNA that has "hitched a ride" on the cellular machinery of mammals: literal genetic parasites that have discovered a way to be replicated generation after generation without ever needing a "body" to do so.

The most abundant gene in the entire *human* genome is reverse transcriptase, a *viral* gene used to make a DNA copy of viral RNA for insertion into the host DNA. It's not the copy of an entire viral genome (Herv) but only one gene of viral origin. It is estimated that over 100,000 copies of reverse transcriptase exist in the human genome, accounting for a staggering 14.6% of the entire genome – 5 times more genetic material than what is clearly

human. Presumably it exists in such staggeringly high quantities because reverse transcriptase has the ability to get itself transcribed and then make a DNA copy of its RNA only to get reinserted into the genome at another location.

Almost as common are shorter sequences of a viral promoter sequence known as Alu repeats. The 280-basepair Alu text may repeat over a million times in the human genome, constituting another 10% of the entire genome. All tolled, more than 35% of human DNA is minisatellite DNA.

The presence of these "genetic freeloaders" has caused many Evolutionary Biologists to rethink evolution itself. Perhaps natural selection has less to do with competition between species or individual organisms within a particular group and is actually far more about competition between genes using individuals and species as mere vehicles for their propagation. Genes may be thus be thought of as "selfish replicators" or "selfish genes," to use Richard Dawkins' terminology.<sup>7</sup> From one perspective, evolution's greatest successes may be those genes that have figured out a way to replicate themselves even though they have dispensed with their bodies altogether. Selfish genes may prove that for some, there really is a "free lunch" in nature.

Unfortunately for their hosts, these genetic parasites are not passive passengers but can pose credible threats to the integrity of the host genome. Replicating in the "wrong place" of the genome could be disastrous, especially given the tendency of selfish DNA to jump from one locus to another seemingly at random. Barbara McClintock discovered this phenomenon of "jumping genes," or transposons, in the 1940s while working with Indian corn, but the importance of her discoveries would not be recognized for many years. It is now estimated that 1 out of 700 spontaneous human mutations are due to the action of "jumping genes" (the number may be as high as 1 out of 10 spontaneous mutations in other mammals).

Mounting evidence suggests a scenario in which viral genes replicate freely within mammalian genomes until the mammalian genome "learns" to suppress the spread of viral genes. Yet the mammalian genome, while able to stop rampant replication of viral genes within its own genome is powerless to excise the genes already incorporated. Thus, we see cumulative evidence of past infections.

The primary means of controlling viral gene replication within the human genome appears to be methylation. A methyl group may be added to a cytosine nucleotide to "shut off" that segment of DNA. Researchers have long thought that methylation of DNA allows for tissue differentiation in an individual cell (nerve cells versus heart or skin cells, etc.), but new evidence suggests that an equally important function of DNA methylation may be to suppress replication of transposons – viral parasitic elements – within the mammalian genome. This may be especially important in the genesis and pathophysiology of cancer since one of the first events associated with cancerous transformation of cells is stripping the DNA of methylation. This allows not only unfettered cell replication, but also unfettered viral transposon replication. This is yet another area where competent methylation capacity is critical for normal physiology.

Ironically, the transposons and minisatellite regions of the DNA were the first to have clinical applications, especially in forensic medicine. These highly repetitive and variable regions of DNA are the source of "DNA fingerprinting" (see discussion below in "diagnostic genetic methodologies"). The minisatellite arrays with the genome can serve as unique identifiers for individuals.

## A Brief Timeline of the Discoveries and Concepts of Heredity and Molecular Genetics

*If I have seen further, it is because I have stood on the shoulders of giants* – Sir Isaac Newton

- 1655 Robert Hooke discovers that living matter is made up of "cells."
- 1759 CF Wolff proposes the general cell theory. 1838 Matthias Schleiden concludes that all plants are made up of cells.
- 1839 Theodor Schwann concludes that all animals are made up of cells.
- 1859 Charles Darwin publishes *The Origin of Species* and discovers the Law of Natural Selection.
- 1865 After breeding ~28,000 pea plants for seven characteristics, Gregor Mendel discovers the Laws of Heredity and publishes *Experiments in Plant Hybridization* in the Brunn Society for the Study of Natural Science's journal. No one noticed; not even Darwin.
- 1869 Johann Miescher isolates DNA from the nuclei of white blood cells as pus in soldier's bandages, but didn't have a clue as to what it was or did.
- 1869 Francis Galton publishes *Hereditary Genius*, claiming that heredity alone is responsible for character traits – his work would evolve into the eugenics movement.
- 1875 Francis Galton demonstrates the usefulness of twin studies in elucidating the relative importance of genes and environment in determining characteristics.
- 1882 Walther Fleming, using dyes to stain cells, discovers and names "chromosomes." 1889 Francis Galton publishes *Natural Inheritance* and describes the quantitative measurement of metric traits in populations. This is the first statistical study of variation, a field now known as biometry.
- 1892 August Weismann proposes that heredity is transmitted by a substance with a "chemical and molecular constitution."
- 1885-1901 Albrecht Kossel discovers that DNA contains the bases adenine, cytosine, guanine, and thymine (ACGT). Nobel Prize 1910.
- 1900 Karl Erich Correns, Hugo de Vries, and Erich Tschermak independently re-discover Mendel's Laws.
- 1900 Karl Landsteiner discovers the blood-agglutination phenomenon in humans, later classified as A, B, O, and AB blood types.
- 1900 Pearson develops the chi-square test in statistics.
- 1902 Archibald Garrod first reports that a human disease (alkoptonuria) behaves as a Mendelian recessive.
- 1902 William Bateson coins the terms, "genetics," "allele," "heterozygote," and "homozygote."
- 1902-3 Walter Sutton formulates the chromosome theory and discovers that they come in pairs. Later he discovered that sperm and egg cells had but one copy of each chromosome.
- 1903 Wilhelm Ludwig Johannsen introduces the ideas of genotype, phenotype, and selection.
- 1908 Godfrey Harold Hardy suggests that Mendelian mechanisms acting alone have no effect on allele frequencies in populations. This observation forms the mathematical basis for population genetics.
- 1909 Wilhelm Johannsen coins the term "gene" for the sections of chromosomes responsible for passing on a particular trait. He also coins the terms, "genotype" and "phenotype." 1909 Archibald Garrod publishes *Inborn Errors of Metabolism*, the earliest discussion of the biochemical genetics.
- 1915 *The Mechanism of Mendelian Heredity* by Morgan, Sturtevant, Bridges, and Muller is published.

- 1920s Thomas Hunt Morgan conducts thousands of experiments with *Drosophila* and shows that genes lie in a row on chromosomes, "like beads on a string." He is awarded the first Nobel Prize for work in Genetics in 1933.
- 1920s Frederick Griffith found that he could kill mice by injecting them with live benign bacteria combined with dead pathogenic bacteria. He postulated a "transforming principle," later shown by Avery in 1944 to be DNA.
- 1927 Hermann Muller discovers that X-rays can induce mutations in *Drosophila*. Nobel Prize 1946.
- 1930s Phoebus Levene discovers that DNA also contains deoxyribose, a sugar.
- 1941 George Beadle and Edward Tatum formulate the one gene/one enzyme theory: each individual gene causes the production of only one protein. Nobel Prize 1958.
- 1944 Oswald Avery proposes that DNA is the stuff of heredity and of genes. He concludes that DNA was the "transforming principle" that killed the mice in Griffith's experiments. Not confirmed until 1952. Until this time, the biochemical mechanisms of heredity were unknown and hotly disputed.
- 1940s Barbara McClintock discovers transposons, a.k.a. jumping genes. Ignored at first, her research became instrumental in explaining "selfish genes." Nobel Prize 1983.
- 1940s Ochoa and Kornberg discover the biosynthetic pathways for RNA and DNA. Nobel Prize 1959.
- 1950 Erwin Chargaff discovers that DNA contains equal numbers of adenine as thymine and cytosine as guanine; known as Chargaff's ratios.
- 1951 Rosalind Franklin uses X-ray diffraction to help elucidate the structure of DNA.
- 1952 Alfred Hershey and Martha Chase prove that DNA is the hereditary material.
- 1952 Frederick Sanger's team works out the complete amino acid sequence for insulin. Nobel Prize in Chemistry 1958.
- 1953 James Watson and Francis Crick discover the double-helix structure of DNA. Noble Prize 1962.
- 1957 Arthur Kornberg synthesizes DNA in a test tube.
- 1959 J. Lejeune, M. Gautier, and R. Turpin show that Down syndrome is trisomy 21, a genetic disorder.
- 1960s Werner Arber, Daniel Nathans and Hamilton Smith discover restriction enzymes in bacteria that cleave viral DNA at specific places. Nobel Prize 1978.
- 1966 Nirenberg and Ochoa "crack the genetic code" by identifying the 3 letter "words" coding for the 20 amino acids.
- 1973 Stanley Cohen and Herbert Boyer create the first transgenic organism by putting a toad gene into a bacteria using restriction enzymes.
- 1978 Louise Brown, the first in vitro fertilized embryo, or "test tube baby," was born.
- 1980 US Patent and Trademark office grants a patent on a genetically engineered bacteria.
- 1982 The first recombinant DNA drug approved for use in humans: insulin from transgenic pigs.
- 1983 Kary Mullis develops polymerase chain reaction (PCR) technology. Noble Prize 1993.
- 1984 Alec Jeffries develops DNA typing (DNA "fingerprinting") after discovering the unique patterning of minisatellite DNA in individuals.
- 1988 US Patent and Trademark office grants a patent for the first genetically altered mammal: the oncomouse.
- 1988 The Human Genome Project begins.
- 1989 Bishop and Varmus awarded Nobel Prize for discovery of the cellular origin of retroviral oncogenes.
- 1990 W. French Anderson performs the first approved gene therapy on a human; Ashanti DeSilva.

- 1994 Calgene Corporation introduces the first genetically engineered food on the market: the FLAVR SAVR<sup>®</sup> tomato.
- 1996 Ian Wilmut clones "Dolly," the sheep from an adult uterine cell, demonstrating the pluripotency of the somatic cell genome.
- 2000 First "draft" of the human genome completed simultaneously by the Human Genome Project and by Celera Corporation.
- 2002 First commercially available laboratory profiles measuring polymorphic variants for predictive genomics for common clinical conditions by Great Smokies Diagnostic Laboratory.

## **Diagnostic Genetic Methodologies**

### ***Genetic Engineering***

The genomics revolution has of course been predicated on the development of new laboratory methodologies that enable us to read the sequence of specific DNA nucleotides, thereby mapping genes themselves. A chance meeting between two scientists at a conference in Hawaii in 1973 set in motion the development of the biotechnology necessary for reading the code of life. Herbert Boyer was a researcher at UCSF working with restriction enzymes. Restriction enzymes are bacterial enzymes whose purpose is to protect the bacteria against viral attack by literally shredding the invading viral DNA into bits. As it turns out, restriction enzymes are absolutely specific in terms of the sequence of DNA they recognize in order to cleave the DNA at precise locations. Moreover, different bacteria express different restriction enzymes recognizing different nucleotide sequences for cutting DNA. Today, more than 1000 different restriction enzymes have been identified. Meanwhile, Stanley Cohen of Stanford was working on bacterial exchange of plasmid DNA and was intensely interested in the potential of using restriction enzymes to facilitate exchange of DNA between bacterial plasmids. Cohen and Boyer began working together to this end.

Their research was simple, elegant, and effective. First, restriction enzymes were used to "snip" or "cut down" sections of plasmid DNA from two different bacteria (one strain resistant to tetracycline, the other resistant to kanamycin) at exactly the same sequence location. Mixing the identically cut down segments of DNA in the presence of a ligase enzyme allowed them to be knit back together and resulted in the formation of new hybrid plasmid DNA. They demonstrated the successful creation of the new hybrids by the transformed bacteria's ability to grow on medium containing both tetracycline and kanamycin. For the first time DNA was transferred directly and *in vitro* from one organism to another by humans. The field of recombinant DNA was born.

Cohen and Boyer asked the next logical question, "Can DNA from different species be as easily exchanged?" Accordingly, they cut down a gene from a toad and mixed it with bacterial DNA. With every new generation of bacteria, the toad gene was present. Species boundaries could be crossed, they demonstrated, and the first transgenic organism was created.

The terms genetic engineering, gene cloning, gene splicing, and recombinant DNA are all used interchangeably to indicate the essential process created by Boyer and Cohen. Today it is common to buy cheese, for instance, that has listed in its ingredients, "microbial enzymes," indicating that the enzyme to make the cheese, originally from cows, has been grown from

transgenic bacteria bioengineered with the bovine enzyme gene. Other areas of transgenic experimentation abound, including novel plant breeding: herbicide and insect resistant plants; nitrogen fixing plants; plants producing vaccines; etc. The list of possibilities is endless because the capacity to mix DNA from different species is endless.

The first commercially marketed, genetically altered food offered to consumers came on the market in 1994. Calgene, a biotechnology company in California, created the FLAVR SAVR<sup>®</sup> tomato: a tomato plant that would not easily bruise or rot once picked off the vine and so could be picked later in the ripening cycle to improve flavor and still make a long journey to the supermarket without being prematurely damaged.

Research into transgenic animals (known also as chimeras) is also proceeding rapidly. Animals can be genetically engineered to produce biological products that humans need. Transgenic pigs, for instance, now produce human insulin for diabetic patients. This type of process - putting human genes into animal carriers - is known colloquially as molecular farming, or "pharming." Further, transgenic genes have been spliced into animals and plants, inducing an adaptive advantage for specifically designed ends. For instance, human growth hormone genes have been spliced into salmon, chickens, and pigs in an effort to produce faster growing and bigger animals for human consumption. Humans' needs for organ transplants may one day be met by growing and harvesting pigs genetically altered to reduce the likelihood of organ rejection.

Critics of bioengineering have coined the term " Frankenfoods " to describe transgenic plants and animals. To be sure, the future of how we use transgenic organisms is yet to be fully determined, but, like it or not, Pandora has opened her box. Since 1986, over 2,000 transgenic plants and nearly as many transgenic animals have been developed and, of course, patented.

## ***DNA Fingerprinting***

In 1984, Alec Jeffreys noticed that certain segments of DNA had short sequences that repeated over and over, sometimes as introns (non-coding segments of DNA that sit in-between the coding segments, known as exons) within a specific gene and sometimes within the larger, non-coding sections of the genome. He called these segments "minisatellites." He also noticed that the pattern of minisatellites was unique to each individual – no two were alike (except, of course, in identical twins). From this he reasoned that minisatellites in DNA could be used as unique identifiers in forensic medicine and elsewhere, much as fingerprints are currently used, only with a higher degree of accuracy.

The procedure for DNA typing is fairly simple: a sample of DNA is extracted and purified and is then cut at specific locations using restriction enzymes, producing DNA fragments of varying lengths. Placing the fragments on a gel plate and passing an electric current across the plate (gel electrophoresis) causes the smaller fragments to migrate more quickly toward the positive pole than the larger fragments. The sorted DNA fragments are then subjected to a blotting technique in which they are split into single strands and transferred to a nylon sheet. A radioactive DNA probe is added which will anneal (bind) to the single-stranded DNA at specific sequences of minisatellite DNA. Finally, a piece of x-ray film is then exposed to the labeled, bound, blotted, and separated DNA to reveal unique patterns of minisatellites.

## ***Polymerase Chain Reaction (PCR)***

DNA fingerprinting only works if you have a fairly large amount of DNA to use in the process. Polymerase Chain Reaction or PCR was developed in 1983 by Kary Mullis in order to generate multiple copies of DNA from a small sample. PCR can take even a single strand of DNA and within several hours generate billions of copies of that DNA. Simply put, PCR allows for exponential amplification of a specific genetic sequence.

PCR is a three-step process carried out in repeated cycles. Ingredients required for the process include, a sample of the DNA or DNA segment to be copied, two short primer sequences that can bind (or anneal) to the template and form a starting point for copying a specific DNA segment, free nucleotides, and DNA polymerase, the enzyme that promotes DNA replication.

First, the DNA is denatured, or separated, by heating it to 95°C (203°F). Second, the temperature is lowered to 55°C (131°F) so that the primers can anneal to the template. Third, the temperature is raised to 72°C (162°F) where DNA polymerase begins adding nucleotides onto the ends of the annealed primer. The entire process takes only about 5 minutes and then can begin again. Usually 20-40 cycles is more than enough to produce an adequate sample of DNA for "fingerprinting" analysis; 35 PCR cycles can produce a little over 17 billion copies of the DNA.

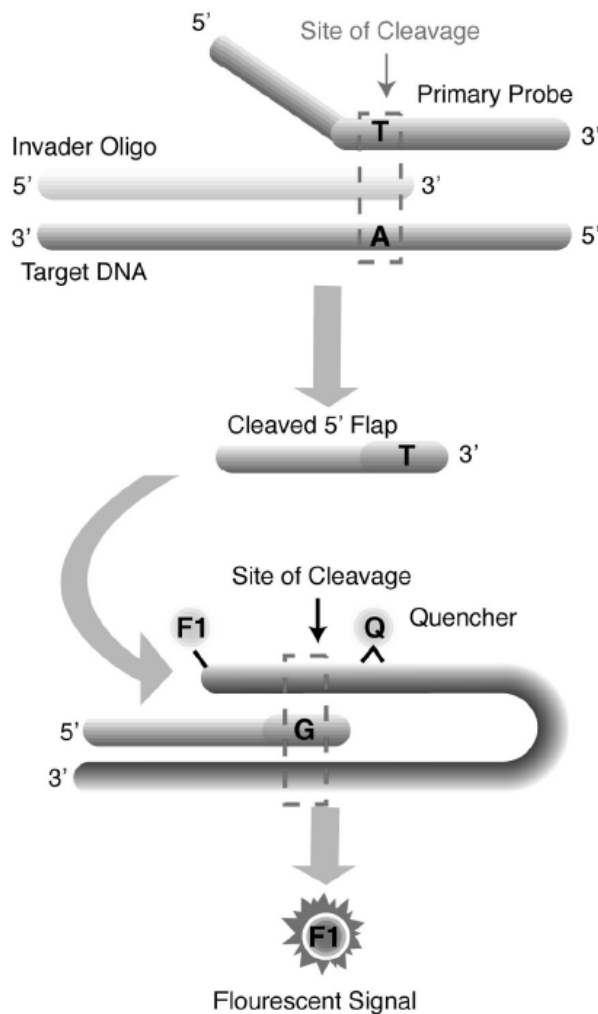
When PCR was first developed it had one serious limitation: new DNA polymerase had to be added after each cycle since the 95°C temperature would destroy human DNA polymerase (it is, after all, a protein). The problem was solved in 1987 with the discovery that a thermophilic bacteria (*Thermus aquaticus*) found naturally in hot springs made DNA polymerase that could withstand the exceedingly high temperatures of their own natural environment (~100°C). By using heat-stable, bacterial DNA polymerase, the cycles can continue uninterrupted and without needing to add new reactants with each cycle. This allowed researchers to develop automated PCR machines, dramatically reducing the cost, time, and skill necessary to carry out PCR.

PCR has many applications since it requires only a very small sample of DNA to be able to work. It is used in forensic medicine when only a small amount of DNA is culled from a crime scene, even a single hair; it is used to detect low-level viral infections like HIV; it is used to amplify DNA fragments found in a 40,000 year old woolly mammoth; and it has been an integral tool in the human genome project, allowing us to map all 3 billion nucleotides.

## ***Invader<sup>®</sup>, Assay***

New, inexpensive DNA assays are coming on the market, making DNA analysis possible for many conditions, including routine single nucleotide polymorphism, or SNP, analysis. A unique diagnostic picture of a person's genetic potential is now commercially available. The Invader<sup>®</sup>, DNA assay is one such method.

Invader<sup>®</sup>, technology is a novel, homogenous platform that can analyze DNA without prior PCR amplification of the target sequence. Instead of amplifying the target itself, the Invader<sup>®</sup>, assay amplifies a target specific signal, detected by a standard fluorescence microtiter plate. So, rather than exponentially amplifying the DNA itself, as in PCR, the Invader<sup>®</sup> platform uses a larger sample of DNA and then linearly amplifies the signal which detects a particular DNA sequence. Invader<sup>®</sup>, technology uses oligonucleotides that bind to the target in a sequence-specific manner and an enzyme that cleaves in a structure-specific manner, enabling it to detect single base changes.<sup>8,9</sup>



The Invader<sup>®</sup>, DNA reaction involves two steps:

In the primary reaction, two synthetic oligonucleotides hybridize in tandem to the target DNA, forming an overlapping structure, or flap. An enzyme recognizes and cleaves off this flap, releasing it as the target specific product. Multiple copies of the primary probe can bind, having its flap clipped off, resulting in the amplification of flaps in the reaction well (see diagram).

In the secondary reaction, the flaps are free to bind to a fluorescence resonance energy transfer (FRET) cassette, creating another overlapping flap this time in the FRET cassette, also cleaved by the enzyme. When the FRET cassette is cleaved, the fluorophore (F) and quencher (Q) are separated, producing a fluorescence signal that can be read.

If the oligonucleotides do not bind, no flap is released, no secondary binding and cleavage takes place and no fluorescence signal is produced or detected. The procedure is sensitive,

accurate, precise, and flexible.<sup>10,11</sup>

### 3. Mendelian Inheritance

(a.k.a. the genetics you probably remember)

The Laws of Heredity are few; their implications for life are vast. The simplest genetic characteristics are those whose presence depends on the genotype at a single locus: one gene controls the expression of one characteristic. Such characters are known as Mendelian. Over 10,000 Mendelian characters have been identified in humans.

After breeding some 28,000 pea plants, Gregor Mendel proposed a number of laws regulating inheritance in plants. He studied seven simple traits (alleles) in pea plants. For instance, some plants were tall, others short; some had wrinkled peas, some smooth, etc. Each of these traits was determined by an allelic variant at an individual gene locus. What we refer to as genes today, he simply called "factors" of inheritance.

In sum, the Laws of Heredity Mendel discovered were,

- I. Each physical characteristic corresponds to a single gene
- II. Genes come in pairs
- III. Only one gene of the pair is passed on to the next generation from the parent
- IV. It is equally probable that either gene will be passed on
- V. V. Some characteristics are "dominant" while others are "recessive"

In spite of a lukewarm reception at the Brunn Society for the Study of Natural Science, Mendel published his findings in the Society's journal.<sup>12</sup> Mendel sent copies of his work to botanists around the world but no one recognized the importance of his findings. Thirty-five years later, and sixteen years after Mendel's death, Correns, de Vries, and von Tschermak each independently arrived at similar conclusions as Mendel had and each was stunned to realize that his own "revolutionary discoveries" were merely repetitions of Mendel's previous work.

A trait (character) is dominant if it is expressed in the heterozygote (only one of the chromosome pair carries the gene) and recessive (both chromosomes carry the gene) if it is only expressed in a homozygote. Dominant and recessive are properties of traits, not of genes themselves. Mendelian pedigree patterns are not always as evident in humans as they are in pea plants due to a number of complicating factors. Chief among these is incomplete penetrance. The penetrance of a character is the probability that a person with the genotype will manifest the dominant character. Other confounders include delayed onset of late-age genetic disorders, multigene effects, and variable expression of genes (different features of a single genetic syndrome will appear in similarly affected individuals). In addition, spontaneous mutations can occur where no pedigree association exists.

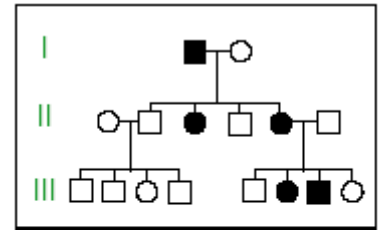
There are six basic Mendelian pedigree patterns:

1. Autosomal dominant
2. Autosomal recessive
3. X-linked recessive
4. X-linked dominant
5. Y-linked
6. Mitochondrial (matrilineal)

Complications aside, the ideal characteristics of these patterns of inheritance are as follows:

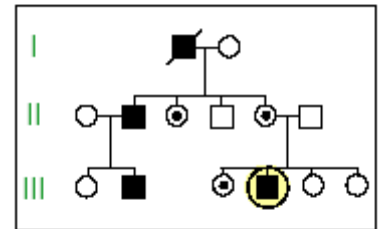
### Autosomal dominant inheritance

- The affected person has at least one affected parent
- Affects either sex equally
- Transmitted by either sex
- Offspring of an affected person have a 50% chance of inheriting the trait
- E.g., Huntington's disease



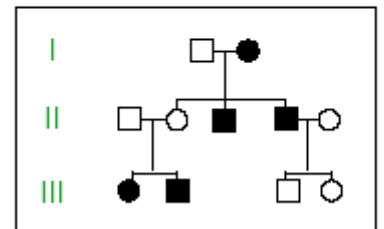
### Autosomal recessive inheritance

- Affected people are usually born to unaffected parents, unless one or both parents is homozygous for the trait
- Heterozygous "carriers" are usually asymptomatic
- Incidence increases with increased consanguinity
- Affects either sex
- The offspring of two heterozygous carriers have a 25% chance of being affected



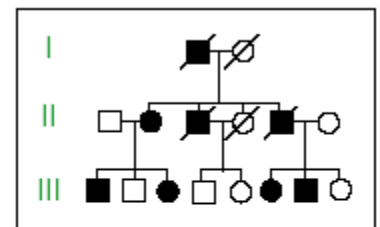
### X-linked recessive

- Affects males predominately
- Parents are usually unaffected (mother is usually an asymptomatic carrier)
- Females can be affected if the father is affected and the mother is a carrier (probability = 50%)
- There is no male to male transmission in the pedigree
- E.g., hemophilia



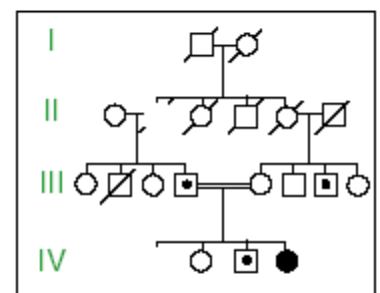
### X-linked dominant inheritance

- Affects either sex but more females than males, since an affected male will always transmit the trait to his daughters
- Females are often affected more mildly than males, and they usually have an unaffected X-chromosome
- The child of an affected female has a 50% chance of being affected regardless of sex
- For an affected male, all of his daughters but none of his sons will be affected



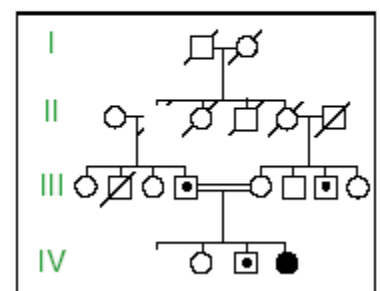
### Y-linked inheritance

- Affects only males
- All sons of an affected father are affected
- Traits are usually not severe
- E.g., hairy ears



### Mitochondrial (matrilineal) inheritance

- Mitochondria are inherited from the maternal egg, not the paternal sperm
- Heteroplasmy (multiple mitochondrial genomes in one individual) is possible but fairly uncommon



Mendelian inheritance patterns were the first evidence to unlock the mysteries of heredity. While 10,000 traits are known to be Mendelian (e.g., eye color), at least as many traits are non-Mendelian. Height, intelligence, personality, and a thousand more characteristics of creatures are multi-factorial, controlled by the interaction of numerous genes, each independently assorted. Furthermore, the same confounders for simple Mendelian inheritance - incomplete penetrance, environmental influences on gene expression, spontaneous mutations - also occur in multi-genic characteristics, but their effects are exponentially multiplied. Still, the Laws of Heredity have taught us much and form the basis from which we learn to know what we do not know about dynamic interactions between genes within the genome and between the genome and the environment.

## 4. Predictive Genomics

### Polymorphic Variation and Single Nucleotide Polymorphisms

Polymorphisms, literally, "many shapes," include any variation in the genome between individuals. What makes you different from your neighbor in genetic terms? As it turns out, it's about 0.0003% of the genome! Polymorphisms can be harmless. If they occur randomly in the DNA, which appears likely, then >97% would be harmless to the individual since they would occur in the non-coding portion of the DNA (recall that only about 3% of the human genome actually codes for genes). But polymorphisms can also be disastrous – cystic fibrosis, muscular dystrophy, sickle cell anemia, and the like are examples of very serious diseases that result from changes in only one letter of the genetic code. Finally, some polymorphisms are somewhere between harmless and full-blown disease – they affect function but only moderately, sometimes improving that function, but more often, impairing it.

In thinking about polymorphisms, it is helpful to remember that there is no final "goal" of evolution; no such thing as evolutionary "progress." Natural selection is simply the process whereby organisms change to fit the limitations and opportunities afforded by a changing physical environment. To the extent that organisms can change to meet their challenges and opportunities, they are successful, i.e., they survive and thrive. Variation within a species improves the likelihood that at least some individuals in that species will be better able to meet those environmental challenges and carry the future of the species within their loins. Species variation at its most basic level occurs in the genome first. Polymorphisms are another name we give to genomic variability within a species.

Landsteiner blood groups (A, B, AB, and O) are an excellent example of polymorphisms and illustrate the evolutionary implications for change. ABO blood groups result from variation in the enzyme galactosyl transferase coded for by 1,062 base pairs within a gene that is ~18,000 base pairs long. There are only seven nucleotides that are different between type A and type B individuals. Three of the mutations are silent, i.e., they code for the same amino acids in the final protein, leaving only four points of functional genetic polymorphism.<sup>13</sup> Further, type O differs from type A by a single deletion: the 258th nucleotide is missing in type O, but this produces enormous changes since there is a frame-shift and every three-letter codon after that is different than for type A.

Is it the case that all blood groups are equally harmless and innocuous and that we are merely observing variation that is neutral from the perspective of natural selection? Was it merely by chance that the Americas were originally populated only with individuals with type O blood? Or is it possible that other blood groups came to the Americas but subsequently died out. Simply put, are there specific adaptive advantages to the various blood groups?

The latter idea appears to have considerable merit when considering the epidemiological evidence. In the 1980s it became apparent that children with certain blood types were more or less susceptible to various types of infections. For instance, type O individuals appear to be more susceptible to cholera; type B are somewhat more resistant; type A more resistant still; and Type AB individuals are virtually immune to cholera. This illustrates a common principle in genetics, known as frequency-dependent selection: adaptive advantage always lies with the rare version of a gene so that neither version can become extinct (recall that if two AB individuals mate, only half of their offspring will be AB; 25% will be BB and 25% will be AA; i.e., if AB individuals survive, so to will BB and AA individuals).

On the other hand, type O individuals are more resistant to malaria and to syphilis, and are less likely to get various types of cancer. Similarly, one in five individuals are unable to secrete ABO blood group proteins into their saliva and other body fluids. These non-secretors are more likely to suffer conditions like meningitis, yeast infections, and urinary tract infections, but less likely to catch influenza or respiratory syncytial virus.<sup>14</sup> This illustrates the general principle that genetic variability often has a lot to do with prevalent infectious diseases. This should not surprise us given our understanding of competition between species as a central element in natural selection: "survival of the fittest."

Other examples abound. Pima Indians have a genetic constitution that would allow them to survive longer than most humans without food, often referred to as the "thrifty genotype." In an age when food supply was anything but certain, this was a distinct adaptive advantage. However, in the current world of ample food, it is a distinct disadvantage. Almost every adult Pima Indian is obese and diabetic. Their "thrifty genotype," once an advantage, is now clearly a liability.

Part of being fittest, from the perspective of the species, is genomic variety. Which might lead us to speculate that the old wives tale of opposites attracting may have some foundation in biology. Indeed, evolutionary behaviorists have supplied us with some intriguing confirmatory evidence. Claus Wederkind and Sandra Furi conducted experiments suggesting that men and women prefer the body odor of members of the opposite sex who are most *unlike* them genetically in terms of major histocompatibility genes that are involved in immunological differentiation between "self" and "not-self." MHC proteins are especially important in recognizing foreign invaders.

Similar findings have been found in mice where females appear to choose mates based on the smell of their urine, choosing mates that are the most dissimilar to them genetically. As an interesting side note, in the human study, only women taking oral contraceptives failed to show a preference for different MHC genotypes. The researchers concluded, "No one smells good to everybody; it depends on who is sniffing whom." And so it appears, "opposites really do attract."<sup>15</sup>

The general principle here appears to be that different versions of genes become more or less frequent in response to the threat of specific diseases, and variation is good for the

species. Yet what is good for the species may not be so good for the individual. Indeed, here we have the crux of the difference between natural history and medicine. Medicine is the art and science of restoring sick *individuals* to health, and has nothing to do with populations or natural selection or adaptive advantages.

From the perspective of medicine, we must view polymorphisms in an entirely different light. In medicine, polymorphic variation is likely to convey greater or less susceptibility toward specific diseases by improving or impairing physiological function. The most common type of polymorphism is known as single nucleotide polymorphisms or SNPs in which, as we have said, a single nucleotide in a gene is changed.<sup>16</sup> Currently, a consortium of private companies and governmental agencies has set for itself the task of identifying and cataloguing as many SNPs as possible and as quickly as possible in order to keep this intellectual property within in the public domain (since genetic variations are patentable under US law). Their goal is to identify 100,000 SNPs in the human genome by the end of 2002.

SNPs analysis may be critical for the complete understanding of complex human diseases since certain genotypes will be consistently associated with those individuals who develop particular diseases – both acute and chronic. Aberrant genes produce aberrant proteins and enzymes. By identifying the genetic aberrations, we may come to a more complete understanding of the molecular basis of diseases, from which novel therapeutics may arise. Thus, population genetics and epidemiological genetics may lead to advances in molecular genetics and to more effective therapeutics for individuals. This research may prove especially fruitful for common chronic diseases like heart disease, diabetes, depression, senile dementia and many cancers, where a combination of multi-factorial genetic and environmental influences are all but certain.<sup>17,18</sup> It is estimated that by the year 2020, 73% of all deaths in industrialized countries will be from non-communicable diseases.<sup>19</sup> The importance of fully understanding chronic multifactorial diseases will become increasingly important for more effective clinical intervention.

To this end, it will also become increasingly important to identify SNPs in individuals that confer greater risk or protection in developing chronic diseases. Those SNPs that will be most important clinically are the SNPs that are relevant to the development of common chronic diseases, have a reasonably high prevalence in the general population, and whose physiological effects are modifiable using diet, nutritional intervention, lifestyle changes, and specific pharmacological intervention. In other words, clinically important SNPs must be relevant, prevalent, and modifiable.

At this stage of our medical knowledge, we do not have a plethora of outcome studies of intervention trials to support specific therapeutic regimens for specific single nucleotide polymorphisms. Some use this limitation as justification for not testing altogether – if we have no proven therapies to ameliorate the adverse physiologic effects of a particular SNP, is it ethical or wise to inform a patient that s/he has a particular genetic limitation?

Even in the absence of clear clinical outcome trials, we can still make sound medical inferences based on our knowledge of biochemistry and physiology. While many therapies have not been validated as specifically effective, many therapies are known to be generally effective for chronic diseases, and we can be absolutely certain that no intervention will definitely be ineffective. Furthermore, it is only by identifying individuals who are at increased risk based on their genetic profiles that we will be able to construct and conduct thorough therapeutic efficacy studies.

Consider the following hypothetical situation: it's difficult to win at poker if you don't know what cards you have been dealt. You don't know which cards to hold, which to discard; you don't know how much to bet or whether to fold and wait for the next hand. By analogy, without some understanding of your genetic strengths and weaknesses, you don't know how to play the genetic "hand" life has dealt you.

Predictive genomic testing is currently available for numerous chronic diseases, including cardiovascular disease, osteoporosis, detoxification defects, and immunological defects associated with the gut associated lymphoid tissue and chronic inflammatory conditions. In each of these areas, functional laboratory testing also exists which allows the practitioner to measure the functional integrity of physiological systems as well as the system's metabolic reserve. The combination of genomic SNP analysis and functional laboratory testing provides a novel, effective, and truly comprehensive assessment of both risk and function.

## **Nature vs. Nurture**

Our health and, indeed, who we are at any given moment in time is the result of a combination of two powerful influences in our lives: our genes and our environment. Virtually all human characteristics, not to mention all human diseases, result from the complex interplay of genetic susceptibility and modifiable environmental factors. Environmental factors should be thought of broadly to include infectious, chemical, physical, nutritional, psychological, and social influences. Literally, "we" are the intersection of our genes and our environment.

Estimating the relative impact of genetic and environmental influences for any particular condition or character is difficult. Many genes express varying degrees of penetrance in the population – you may have the gene that predisposes you to some condition, but the gene never manifests and you never get the disease. Often this variable penetrance is due to environmental influences that affect gene expression. For example, a person may have a defect in n-acetyl transferase but never develop bladder cancer because she did not smoke cigarettes, whereas another person who had the same defect and did smoke would have 7 times the relative risk for developing bladder cancer.<sup>20</sup> In addition, many complex diseases like heart disease or diabetes are multifactorial both in terms of genetic influences and environmental influences. Dozens of genes and even more environmental "exposures" may play a significant role in pathophysiology.

How do researchers determine the relative contribution of environment and specific genes in determining the physical manifestation of a condition or character (phenotype)? Twin studies have long been considered the gold standard since the larger the genetic component of a condition, the higher the expected congruence between identical twins. By contrast, the greater the environmental influence the less the expected difference between congruence between identical twins and congruence between fraternal twins. Indeed if fraternal twins have the same frequency of congruence of a condition as identical twins do, it is assumed that the condition is largely environmental (but not entirely environmental, for fraternal twins still share half their genes in common as do any pair of siblings). Twin studies have allowed researchers to conclude that cancer of the stomach, colon, lung, breast, and prostate have a heritability of between 26 and 42%, suggesting that while environment plays a larger role in these cancers, genetic predisposition plays a nonetheless significant role.<sup>21</sup>

In recent years, with the advent of feasible genetic testing for single nucleotide and other polymorphisms, large epidemiological studies have begun to replace twin studies as the preferred method for estimating genetic and environmental influences in diseases. Starting with diseased individuals, scientists begin looking for SNPs within that population that appear in higher frequency than in healthy populations. Candidate genes, once identified, may then be followed in prospective clinical trials to observe if carriers manifest a disease more frequently than noncarriers. Multiple regression analysis can also allow an estimation of the relative influence of numerous genes in the development of a single complex chronic disease.

From the environmental side of the equation, other epidemiological studies can provide useful information. For instance, rates of breast cancer among Asian- American women in the United States vary widely depending on how long they have been in America. First generation immigrants have a breast cancer risk that is similar to their native homelands and is 80% lower than the cancer rates for third-generation Asian-American women. Since both populations have similar genes, the huge difference in breast cancer rates is likely largely due to environmental differences of their new homeland: altered diets, toxin exposure, lifestyle, etc.<sup>22</sup>

The task for any given chronic disease is Herculean, and while new evidence is being published every day, it will likely be some years before we get precise estimates of the relative contribution of specific genes and specific environmental factors in every chronic disease. Yet with each new polymorphism identified and with each new epidemiological study and intervention trial we come closer to a full appreciation of the complex pathophysiology of chronic disease processes. One thing seems clear already, however: in virtually every disease, both genetic susceptibility and environmental influences play pivotal roles in disease and character development. It is almost never a case of Nature or Nurture, and almost always a case of Nature *and* Nurture. This is comforting, for while we may not be able to alter our genes, we can alter our environment.

## **5. Case Study Examples**

### ***Cardiovascular Disease Refractory to Treatment***

Take a hypothetical patient at increased risk for developing cardiovascular disease, for instance. Physical exam and a comprehensive laboratory cardiovascular risk profile reveals that the patient has hypertension, low HDL cholesterol, and elevated homocysteine levels, each of which would increase the risk of developing cardiovascular disease. You prescribe additional magnesium, folic acid, B12, B6, and aerobic exercise. Three months later you reevaluate only to find that none of the parameters has changed.

You decide to investigate further and order a cardiovascular predictive genomics profile that identifies polymorphisms in three critical enzymes in your patient:

- 1) methylenetetrahydrofolate reductase (MTHFR)
- 2) cholesteryl ester transfer protein (CETP)
- 3) angiotensin receptor (AGTR1)

Elevated serum homocysteine indicates a defect in methylation capacity and is independently associated with a 3-fold increase in the risk of a cardiovascular event. In most patients, homocysteine levels fall when the patient is supplemented with folic acid, B12, and B6, but in ~28% of patients, this therapy is ineffective because they have a single nucleotide polymorphism (677C→T) in a critical enzyme that regenerates methionine from homocysteine: methylenetetrahydrofolate reductase.<sup>23</sup> With this polymorphism, an individual cannot efficiently convert 5,10-methylenetetrahydrofolate into 5-methyl tetrahydrofolate, the active methyl donor, and methylation slows. Folic acid is ineffective as a treatment. However, 5-methyl tetrahydrofolate, the "downstream" product of the impaired biochemical pathway, can be supplemented directly. Giving this special, albeit more expensive, form of folate, improves methylation capacity, lowers the homocysteine levels, and reduces CVD risk. In addition, supplementing with betaine engages a secondary "back-up" pathway that also promotes efficient remethylation of homocysteine.

High-density lipoprotein (HDL) cholesterol concentration is inversely related to the risk of coronary artery disease. The cholesteryl ester transfer protein (CETP) has a central role in the metabolism of this lipoprotein and may therefore alter the susceptibility to atherosclerosis. Your patient has B1B1 genotype for CETP (Taq1B; intron 1) that is associated with higher CETP levels and 30% lower HDL cholesterol levels.<sup>24</sup>

As for treatment options, subsequent studies have shown that moderate alcohol consumers with the B1B1 genotype had 30% lower CETP activity and 48% higher HDL cholesterol levels than those with the same genotype who consumed no alcohol.<sup>25</sup> Furthermore, B1B1 genotypes for CETP have been shown to have dramatic reductions in cholesterol levels from taking Pravastatin, one of the statin drugs.<sup>26</sup> Natural statin mimetics like red rice yeast might reasonably be considered as an alternative since the side effects are considerably less than pharmaceutical statin drugs.

Angiotensin II is an important effector controlling blood pressure and volume in the cardiovascular system. Its importance is reflected by the efficacy of angiotensin-converting enzyme inhibitors in the treatment of hypertension and congestive heart failure. Type 1 receptors mediate the major cardiovascular effects of angiotensin II. One SNP of the AGTR1 gene (1166A→C) has been strongly associated with hypertension. AGTR1 antagonists like losartan (Hyzaar®) are likely to produce the best clinical response in patients with receptor polymorphisms.<sup>27,28</sup> Likewise ACE-inhibitors would likely be less effective in this individual, since the dysfunction is at the level of the receptor and not with angiotension conversion itself. This illustrates one of the potentially clinically useful aspects of pharmacogenetics – finding the right drug for the right individual based on his or her genetic polymorphic limitations.

In this hypothetical example, we see how genotyping of an individual's SNPs can give enormous direction in terms of therapeutic intervention: isolating the right therapies for the right individuals. Through the judicious use of diet, nutrition, lifestyle changes and pharmacological therapies it is possible to modify the expression of the genes and to overcome genetic limitations of biochemical pathways. Predictive genomic testing allows us to be smarter clinicians with more effective therapeutics and fewer side-effects. Furthermore, the therapeutic gains are measurable through continued functional laboratory testing.

The genomic revolution is happening now. Medicine will never be the same. Truly individualized medicine is rapidly becoming a reality for us as practitioners and for our patients, who only stand to gain from our increased diagnostic genomic capabilities.

## 6. A Few Bioethical Considerations

With every new paradigm shift in medicine, ethical issues arise, and genomic testing is no different. A plethora of bioethical and social issues arises in the face of genetic testing, especially if the genetic testing reveals a condition or conditions for which no medical treatment is currently available. In Sophocles' tragedy "Oedipus Rex," Teiresias, the blind seer of Thebes, has been given the power to see the future but cannot change it. Perceiving Fate is a heavy burden, indeed. He says to Oedipus, "It is but sorrow to be wise when wisdom profits not."

In the case of many Mendelian diseases like Huntington's disease, cystic fibrosis, sickle cell anemia, and Tay Sachs disease, to name but a few, knowledge of an almost inevitable fate carries with it significant psychological and emotional burdens, not just for the affected individual, but often also for all members of the extended family. All genetic conditions are, by definition, familial. Diagnosis of a genetic disease in one individual raises the specter at least of diagnosis in other members of the family, or confirmation that other family members are carriers and will put any future offspring at risk as well.

Not surprisingly, the penetrance of genetic conditions runs the gamut from nearly 100% likelihood of developing the condition as in Huntington's disease and a 40- 80% chance of developing Alzheimer's disease given an apoE4 allele to a 10-30% chance of developing heart disease with a defective MTHFR allele. But in reality, there are extremely few conditions for which penetrance nears 100%. If penetrance is anything less than 100%, phenotypic expression is mediated by two important factors. First, most complex conditions have multiple gene interactions. There is not one gene locus that determines heart disease or cancer. Dozens of genes are likely to modify risk and ultimately play a role in phenotype expression. Second, the expression of any given gene is modified by environmental factors including diet, lifestyle, toxic exposure, etc. Invariably our health is a combination of nature AND nurture, or what is sometimes referred to as gene-gene-environment interaction.

This is especially true of predictive genomic testing where we are attempting to estimate and reduce the risk of developing complex chronic diseases. The focus is on relative risk given a specific polymorphism and on modulating that risk using environmental modifications. It is useful to recall the four criteria for clinically useful predictive genomic testing.

**Relevant-** The polymorphisms identified should exert significant influence on the development of disease. Polymorphisms should be carefully selected based on their direct influence over specific biochemical pathways that create known symptom clusters or diseases.

**Prevalent-** Polymorphisms should have a significant prevalence in the population so that testing is practical and economically feasible, and so that outcomes trials are likely to occur and novel therapeutics developed and validated.

**Modifiable-** Polymorphisms for which specific risk reduction strategies are known, including dietary, nutritional, lifestyle, and pharmaceutical interventions, are the most clinically useful and pose few bioethical issues.

**Measurable-** Functional laboratory testing should be available to measure risk reduction as well as to monitor therapeutic progress, by evaluating phenotypic expression of genetic tendency, functional integrity, and metabolic reserve.

Fortunately, in predictive genomic testing, since practical intervention strategies are generally available, genetic diagnosis will likely do far more to relieve stress rather than to increase it. Furthermore, since phenotypic or physiologic progress may be monitored using functional laboratory testing, we can monitor patients closely, changing therapeutic regimens as appropriate. In predictive clinical genomics, rather than being a harbinger of a fate to come, SNP testing may be the first step towards comprehensive risk reduction or comprehensive treatment strategy.

## 7. Resources

There are a plethora of websites springing up on the Internet that deal with genomics – even the entire human genome, all 3 billion letters, is available online (talk about your somnolent reading). Simply searching with the word "genome" will elicit hundreds of sites. Obviously, some are more user-friendly than others.

One of the best websites, especially focused on single nucleotide polymorphisms is a site run by the National Center for Biotechnology Information: Online Mendelian Inheritance in Man (OMIM) at <http://www.ncbi.nlm.nih.gov/Omim/>. It is extraordinarily useful in researching individual SNPs and the enzymes and other proteins for which they code. References are cross-linked to PubMed so that abstracts of relevant studies may be reviewed instantly.

Another very useful site is the Office of Genomics and Disease Prevention, run by the Centers for Disease Control, with numerous Powerpoint slide shows available on a wide variety of topics. The site is best located by entering "Office of Genomics and Disease" in a commercial search engine.

An excellent website for vetting bioethical concerns may be accessed at the Genetics and Ethics page, <http://www.ethics.ubc.ca/brynw/>

## Footnotes

1. Bloom BR. Genetics and developing countries. *BMJ* 2001;322:1006-7.
2. Donnai D, Elles R. Integrated regional genetic services: current and future provision. *BMJ* 2001;322:1048-52.
3. Ridley M. *Genome: an Autobiography of a Species in 23 Chapters*. New York: Perennial, 1999.
4. Campbell J. *Grammatical man: information, entropy, language and life*. London: Allen Lane, 1983.
5. Mitchell W. "Food and the Naturopathic Physician." In MJ Cronin, ed., *The Best of Naturopathic Medicine*. Tempe, AZ: Southwest College Press, 1996.
6. Gesteland RF, Atkins JF, eds. *The RNA World*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1993.
7. Dawkins, R. *The Selfish Gene*. Oxford, UK: Oxford University Press, 1990.
8. Lyamichev V, Mast AL, Hall JG, et al. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol* 1999;17:292-296.
9. Lyamichev VI, Kaiser MW, Lyamicheva NE, et al. Experimental and theoretical analysis of the invasive signal amplification reaction. *Biochemistry* 2000;39:9523-9532.
10. Kwiatkowski RW, Lyamichev V, de Arruda M, Neri B. Clinical, genetic, and pharmacogenetics applications of the Invader assay. *Mol Diagn* 1999;4:353-364.
11. Eis PS, Olson MC, Takova T, et al. An invasive cleavage assay for direct quantitation of specific RNAs. *Nat Biotechnol* 2001;19:673-676.
12. Mendel G. Versuche über Pflanzen-Hybriden. *Vorgelegt in den Sitzungen vom 8. Februar und 8. März 1865*
13. Yamamoto F, Clausen H, White T, et al. Molecular genetic basis of the histo-blood group ABO system. *Nature* 1990 May 17;345(6272):229-33.
14. Hill AVS. Genetics of infectious disease resistance. *Curr Opin in Genetics and Develop* 1996;6:348-353.
15. Wedekind C, Furi S. Body odor preferences in men and women: do they aim for specific MHC combinations or simple heterogeneity? *Proc R Soc Lond B Biol Sci* 1997;264:1471-1479.
16. Kwok PY, Deng Q, Zakeri H, et al. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* 1996;31:123-126.
17. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516-1517.
18. Kwok P, Gu Z. Single nucleotide polymorphism libraries: why and how are we building them? *Mol Med Today* 1999;5:538-543.
19. Lopez AD, Murray C. The Global Burden of Disease. Report for the World Health Organization at <http://www.hsph.harvard.edu/organizations/bdu/summary.html>
20. Taylor JA, Umbach DM, Stephens E, et al. The role of N-acetylation polymorphisms in smoking-associated bladder cancer: evidence of a gene-gene-exposure three-way interaction. *Cancer Res* 1998 Aug 15;58(16):3603-10.
21. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000 Jul 13;343(2):78-85.
22. Ziegler RG, Hoover RN, Pike MC, et al. Migration patterns and breast cancer risk in Asian-American women. *J Natl Cancer Inst* 1993 Nov 17;85(22):1819-27.
23. Engbersen AM, Franken DG, Boers GH, et al. Thermolabile 5,10-methylenetetrahydrofolate reductase as a cause of mild hyperhomocysteinemia. *Am J Hum Genet* 1995 Jan;56(1):142-50.
24. Freeman DJ, Griffin BA, Holmes AP, et al. Regulation of plasma HDL cholesterol and subfraction distribution by genetic and environmental factors: associations between the TaqI B RFLP in the CETP gene and smoking and obesity. *Arterioscler Throm*. 1994;14:336-344.
25. Hannuksela ML, Liinamaa MJ, Kesaniemi YA, Savolainen MJ. Relation of polymorphisms in the cholesteryl ester transfer protein gene to transfer protein activity and plasma lipoprotein levels in alcohol drinkers. *Atherosclerosis* 1994 Sep 30;110(1):35-44.
26. Kuivenhoven JA. The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. *New Eng J Med* 1998;338: 86-93.
27. Miller JA, Thai K, Scholey JW. Angiotensin II type 1 receptor gene polymorphism predicts response to losartan and angiotensin II. *Kidney Int* 1999 Dec;56(6):2173-80.
28. Azuma H, Niimi Y, Hamasaki H. Prevention of intimal thickening after endothelial removal by a nonpeptide angiotensin II receptor antagonist, losartan. *Br J Pharmacol* 1992 Jul;106(3):665-71.