S. H. Wilson / Californian Journal of Health Promotion 2007, Volume 5, Special Issue (Hlth Disparities & Soc Justice), 147-163

Genes, Environment, and Health Disparities: Risks and Benefits of Gene-Environment Interactions Research

Samuel H. Wilson

National Institute of Environmental Health Sciences

This is a transcript of a presentation, *Genes, Environment, and Health Disparities: Risks and Benefits of Gene-Environment Interactions Research*, presented at the 4th Annual Summer Workshop, Disparities in Health in America: Working Toward Social Justice, held at The University of Texas MD Anderson Cancer Center, Houston, Texas, June 24-30, 2006.

In this presentation, I want to jump back and forth between the topic of human genetics and the notion that gene-environment interactions control our responses to stressors in the environment, so far as this relates to human health and disease. I also want to spend time at the end discussing my views on the risks and benefits of information from this kind of research. With its profound implications for health disparities, I think this research is especially important for this group, for this symposium, and for the nation, and I very much applaud the organizers of the symposium for addressing the subject in such a robust and effective way.

As we can tell from cover stories in several national publications and other widespread media coverage, genetics and genomics are here, they're here to stay, and they are certainly in the public's consciousness. It seems as though every week there is a new headline trumpeting the identification of a new gene – genes for breast cancer, for intelligence, for just about every human disease or quality. While some of that publicity is exaggerated, the media coverage is useful in informing us that a genetic age is upon us. We've made a remarkable amount of progress in recent years, but human genetics is still a new area, a topic with a whole new set of definitions, and a topic with a whole new set of policy and health disparity concerns.

Let's first define some of our terms. What is a gene? A gene is the functional or physical unit of heredity within our DNA. All living organisms, of course, have genes and have DNA responsible for heredity. Genetics is the study of the science of heredity, and it has been around for a long time. Genome refers to the complete DNA sequence within an organism, that is, every single base pair for the entire unit of heredity. Genomics is actually a new field in biomedical research, developed just recently, and founded on advances in discoveries of the actual sequences of genomic DNA in various organisms. And, the filed of genomics is making use of this information in a variety of ways. Genomics is the study of complete sets of genes – their expression, output, and interactions. What one often can see in the scientific literature nowadays is the five-letter term, "-omics" that refers to analysis making use of information from the entire or whole genome. We now have fields called proteomics, metabolomics, transcriptomics, and many more –omics sciences. These analyses extend to DNA, to proteins, and to all sorts of components within cells, and are referred to by this blanket term, "-omics" Sciences.

Let's look at some of the key milestones for how the history of molecular genetics fits together and how it emerged into where we stand today – involving a huge increase in the size of the DNA databases (Figure 1).

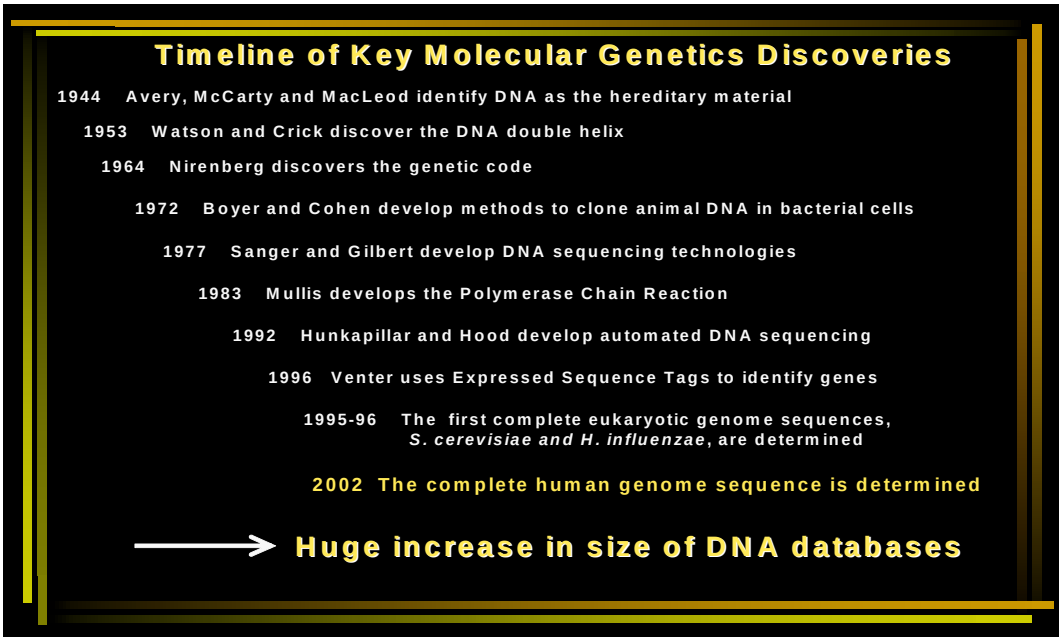

Figure 1
Timeline of key molecular genetics discoveries.

The development of today's genomics really began with the original discovery in 1944 that DNA is the hereditary material. In 1953, there's the famous Watson/Crick double helix discovery, and in 1964, Marshall Nirenberg and his colleagues working at the intramural program at the NIH discovered the genetic code, along with several other laboratories. The next four items in Figure 1 are basically technology development – technologies that enabled us to understand the information that we're going to look at in more detail shortly. The discovery of expressed sequence tags in 1996 by Venter actually drove the sequencing of the human genome. And then in 1995 and 1996, we saw the first examples of the complete sequencing of a eukaryotic genome (that is, an organism with a nucleus) – that was in yeast. This sequencing of the yeast genome produced a tremendous explosion of discovery throughout biology, with the subsequent sequencing of many other organism's genomes. It led many people in the field to believe that if we could just get the human genome sequenced, then we would be able to make fundamental advances in human health and disease. In 2002, as you are probably aware, this was achieved, with the report of a draft version of the human genome. Now we have a much more refined sequence of an example of the human genome.

This sequencing explosion has led to the accumulation of a vast amount of information in DNA databases. Figure 2 shows an example. Up to 2001, as you can see, there was an exponential increase in DNA base pair and gene sequence information in the database. Many of us in biomedical research were successful in cloning genes by the now old-fashioned techniques in the early to mid-1980s. But, you can see that at that point in time, even though we thought we already had a lot of sequence information, it was almost nothing on the scale of what we have today. This chart only displays the growth in information up to 2001, and of course the curve has gone up tremendously since then.

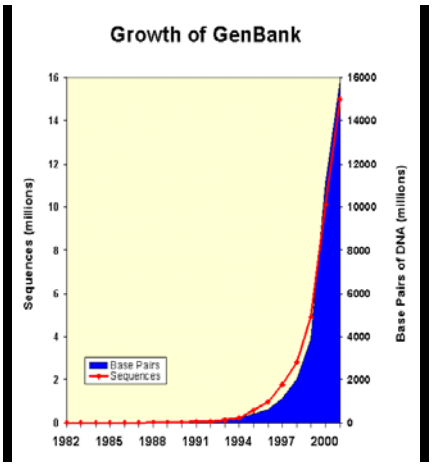

Figure 2
Growth of GenBank

What can we do with this DNA sequence information? Here's one example (Figure 3) for the human genome.

Figure 3 shows the identification of the gene types for the various genes in the human genome. Looking at this pie chart, you can see some of the identified components, such as the nucleic acid enzymes, the DNA polymerases, the RNA polymerases, DNA ligases, and so on. The pie chart also is good news for all of us in medical research: It means that the molecular function of ~36% of the genome is yet to be discovered.

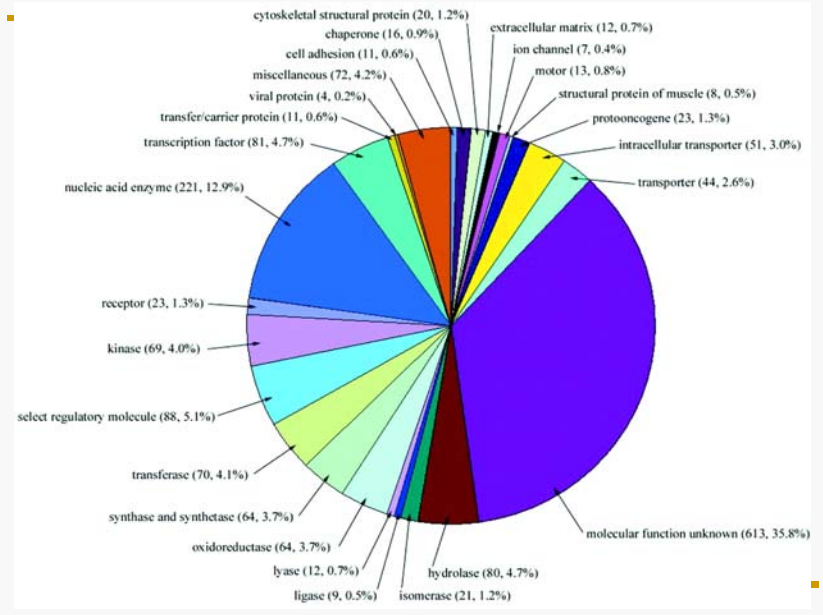

Figure 3
Identification of gene types in the human genome

As you can see in Figure 4, the availability of DNA sequence information has stimulated a huge amount of new activity in biomedical science. We now have functional genomics, the study of what DNA actually does in the cell, human polymorphism studies or the study of genetic variations, comparative genomics, where we study genomic differences between species to shed light on human characteristics, informatics, the use of high-powered computing and statistical analysis to mine useful information from the DNA sequences, and evolutionary genomics, where we try to link different origins in the evolutionary trees. Further down, you can see that the DNA sequence databases also contribute directly to global expression studies, which we'll look at in more detail shortly. These studies are the basis of the so-called "-omics" sciences such as proteomics, which includes the search for biomarkers of disease through studies of protein expression. Experiments and assays in the -omics sciences often make use of very high throughput technology equipment such as DNA microarrays, which can run literally thousands of automated experiments on a single plate. All of these new sciences are developing quite rapidly; they all tie in together at some level, and it won't be long before we see their impact entering routine medical practice.

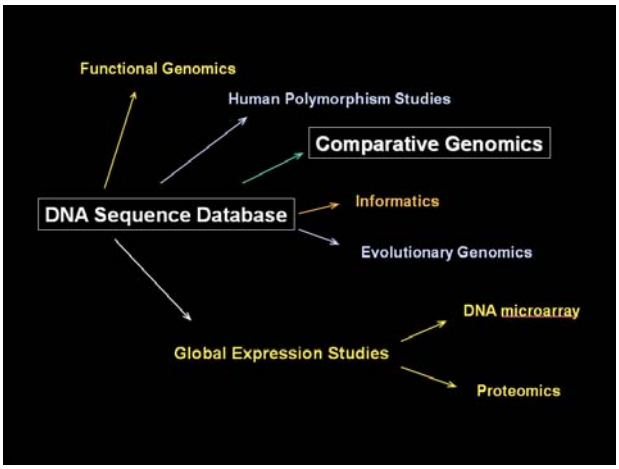

Figure 4
Availability of DNA sequence information has stimulated new activity in biomedical science

Let's look more closely at comparative genomics. Venter et al. published an excellent definition of this important field in Science in 2001. They defined comparative genomics to be the "distribution of the molecular functions in "orthologs" conserved between humans and all other life forms – orthologs being genes that evolved from the same ancestral locus. This simply means that the same type of gene occurs in other life forms, so therefore one can study, for example, bacterial viruses, bacteria, or plants, or other forms of animals, in order to identify the genes and understand what their function might be in higher organisms such as humans.

Figure 5 gives us a picture of what we know about the numbers of genes in different organisms. In the human genome, we have about 20,000 different genes. As I mentioned earlier, 36% of the human genome is not yet characterized with regard to gene content, so this number of genes (20,000) is a moving target. In the mouse genome, it's about the same situation.

The yeast genome has about 6,000 genes. *E. coli*, the classical molecular biology tool, has about 4,000 genes. And of course, there are the viruses that we've used so productively in biomedical research. For example, T4, probably the most famous of these, has 200 genes. Influenza viruses, all of these really small viruses known as RNA viruses, have a very limited number of genes – only 12. These viruses actually make use of the cellular genome in order to achieve replication. What this all means is that in comparative genomics, we can look at yeast and some of these other simple organisms and understand functional aspects that apply to humans, and also to experimental animals.

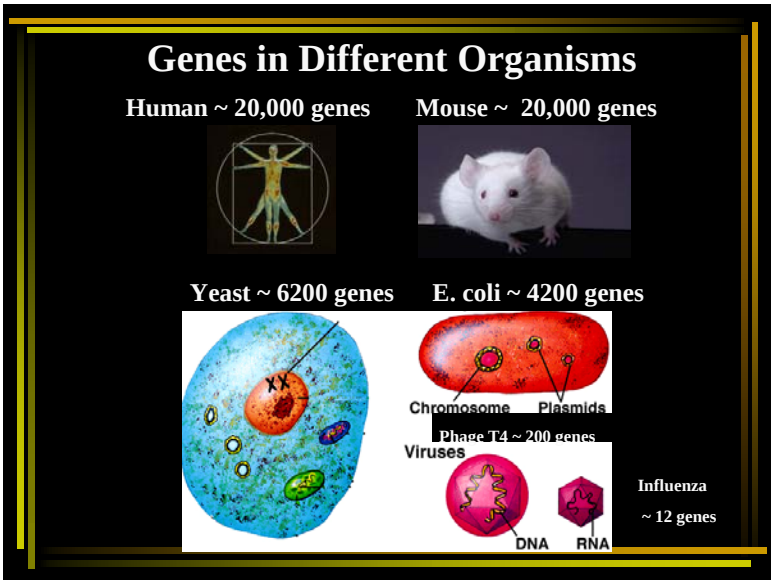

Figure 5
Numbers of genes in different organisms

Looking back at Figure 4, I want to return in more depth to this topic of human polymorphisms, because this is the essence of what I want to talk about and focus on today. To define what a polymorphism is, brings us to a really fundamental question: "Is genomics important for understanding human health and disease?" I can tell you straight away that "yes, it is very, important."

Here is one of the reasons: Genomics allows us to attempt to answer the question, "Why does one person get a disease while another person remains healthy?" We all see our neighbors who smoke ten packs of cigarettes a day and never have any kind of health impact from it, and then others who smoke just rarely who come down with lung cancer or other kinds of health impacts. Clearly, there is individual susceptibility at play as to whether or not we get sick, and this relates fundamentally to these polymorphisms, or genetic variations among individuals.

Individuals differ genetically, and this largely is manifested in these DNA polymorphisms. Individuals have different behaviors, of course, and consume different foods, and are exposed to different toxicants and different forms of stress from the environment. This complex, variable array of environmental factors combines in various ways with complex, variable genetic factors to cause every common disease. That's what we call gene-environment interaction. And

these individual differences in environmental exposures and genes also result in differences in individual susceptibility to disease. So the person who smokes a great deal but remains healthy and the person who smokes rarely but gets lung cancer have different susceptibilities to the effects of exposure to cigarette smoke.

There are differences in the genes that control our responses to environmental factors. These are DNA variations, variations in the sequence of DNA, comparing one person to the next. The variations are largely (but not exclusively) called SNPs, which is our shorthand for single nucleotide polymorphism. Let's look at what these SNPs are, and why they're so vitally important to our understanding of human health and disease.

A SNP is a nucleotide position that varies from one person to another. In the analogy we often use, where the gene is the book of life, a SNP a spelling error, a typo. It's a one-letter variation from person to person. Sometimes they're harmless, sometimes they're helpful, and sometimes they're extremely damaging.

The human genome has, on average, one SNP per 1,000 base pairs (to be defined below). Most of our relatively small genes are in the size range of 30,000-40,000 base pairs. So, at one SNP per 1,000 base pairs, we would have 30-40 SNPs present, even in a relatively small gene. Multiply that number by the thousands of genes in the human genome, and you can start to appreciate that we differ from one to another in DNA sequence by quite a bit, by virtue of these single nucleotide polymorphisms or SNPs.

Genomics allows us to study how individual gene variation or polymorphism influences susceptibility to disease. Let's look in a little bit more detail now by taking a short primer on what these SNPs really are, because I want you to understand these concepts of the genetic code, how we can all be very active and vigorous human beings, but at the same time have differences in our genetic code, one individual to the next.

This cartoon (Figure 6) emphasizes that in the nucleus of our cells, we have the genomic DNA, which acts as the instruction book for how the cell performs, and in complex organisms like human beings, for how the body performs. DNA is illustrated here as the famous Watson-Crick double helix with each nucleotide base paired with its complementary base, A opposite T and G opposite C. These combinations are called "base pairs." So how does all of this information flow work?

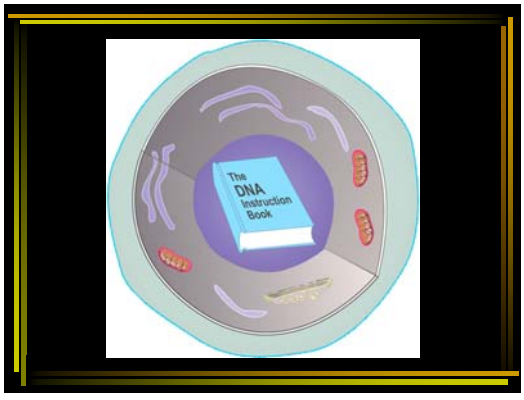

Figure 6
A cartoon reflecting the genomic DNA

As depicted in Figure 7, the DNA instruction book sends information from the nucleus out into the cytoplasm of the cell where the function takes place, as seen in the yellow area of the illustration. This happens through what we call messenger RNA, which transfers the information to the cytoplasm. The messenger RNA is then translated into proteins that are activated and perform functions in the cell. Of course, some of these proteins leak out of the cell and float around in the blood and do various things, and some of them actually talk to cells that are right next to this cell, in order to control growth and differentiation in tissues.

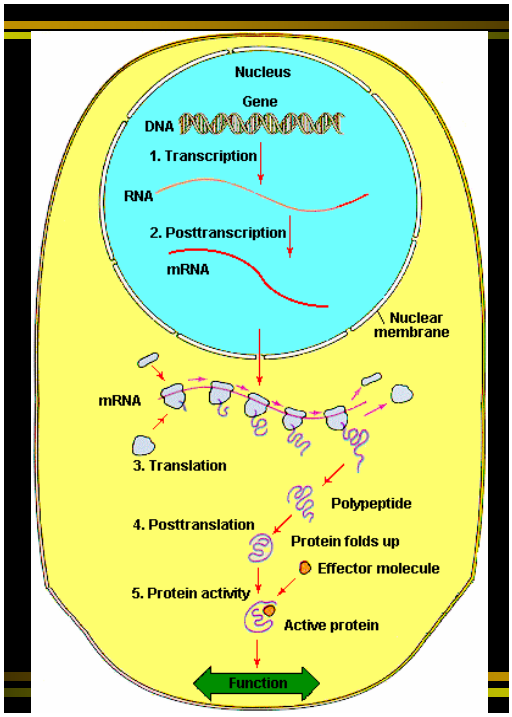

Figure 7
DNA instruction book sends information from the nucleus out into the cytoplasm

Here's another model of the same thing. What we're seeing in green here in Figure 8 is the cell nucleus with the DNA inside, all wrapped up into packages called chromosomes. If you were to take a pair of tweezers and stretch out the DNA, it would be a long, linear molecule of base pairs. DNA is packaged up into chromatin, and into individual chromosomes. You've all seen this illustration of a chromosome or heard about it. But again, lets picture going into the chromosome and grasping the DNA with tweezers and then stretching out the DNA completely. You can see that we come out with these pairs of two nucleotides on opposite strands, represented by the letters, A, C, G, and T. A pairs with T, and C pairs with G – this is termed a base pair.

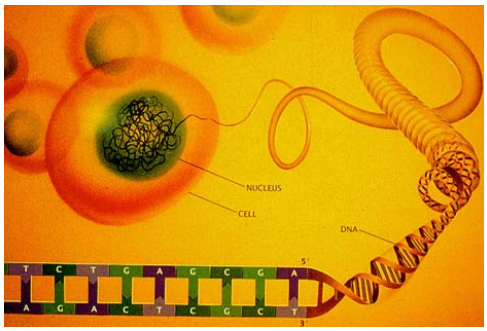

Figure 8
The cell nucleus with DNS inside, all wrapped into packages called chromosomes

Figure 9 illustrates an example of a single nucleotide polymorphism, or SNP on one strand of DNA. Let's say on the left we have the common situation on a chromosome, where the sequence top to bottom is CGACT (note that for simplicity, this image does not include that these bases are paired with the complementary base on the other strand of DNA)., But then, in some human beings, this C shown in the box is changed to G, so that the sequence now is CGAGT, as shown on the right-hand side. This is what we're talking about when we use this term single nucleotide polymorphism or SNP. Again, why is this important? Let's look at a few more simple cartoons to see why.

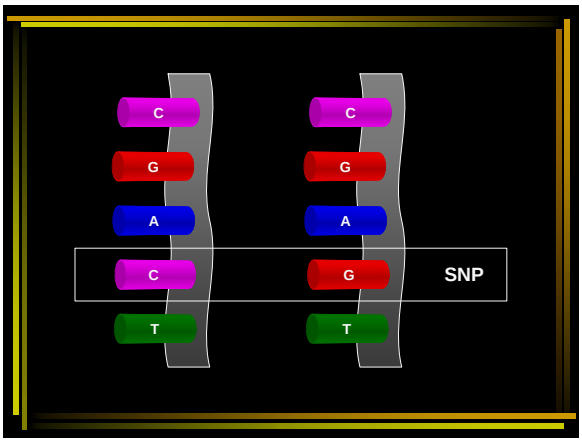

Figure 9
An example of a single nucleotide polymorphism, or SNP on one strand of DNA

Remember that I told you that the DNA instructions are transcribed into messenger RNA. Here's an example (Figure 10) where the DNA actually contains the genetic information, right here in this cart. As you can see, the messenger RNA has a code. In this case, it's

ACGT. The messenger then transfers the code, the ACGT, to the protein factory in the cytoplasm (Figure 11), and then the protein is made (Figure 12). Generally, the protein is engineered to be very strong and active.

But what happens if we have a polymorphism, a SNP, so that the sequence is CCGT instead of ACGT (Figure 13)? In this case, the messenger doesn't really know that there's anything askew here, and takes the message into the protein factory (Figure 14). And lo and behold, what happens is we get a puny protein out on the other side (Figure 15). The small puny protein is not able to accomplish its function as well as the big strong protein, and eventually the cell suffers, especially if stressed. This is the idea of the significance of the single nucleotide polymorphism.

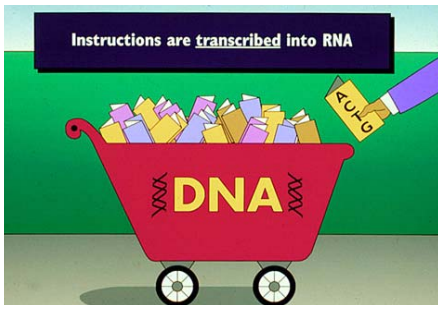

Figure 10
DNA actually contains the genetic information

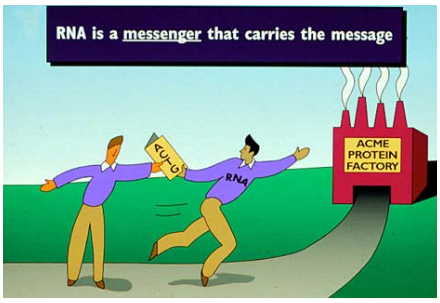

Figure 11
The messenger RNA then transfers the code, the ACGT, to the protein factory in the cytoplasm

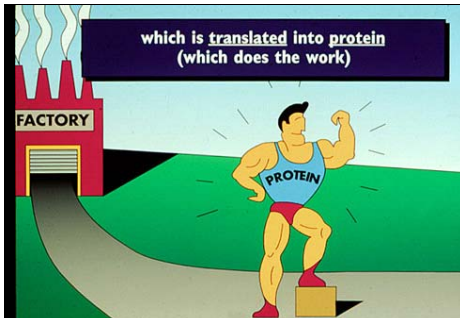

Figure 12
Protein is made

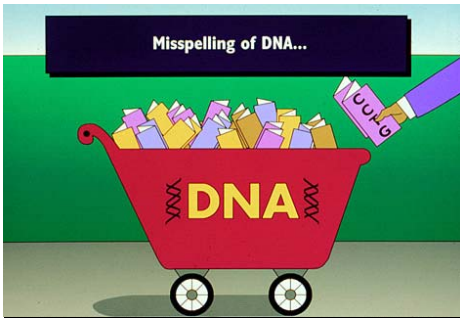

Figure 13
Misspelling of DNA

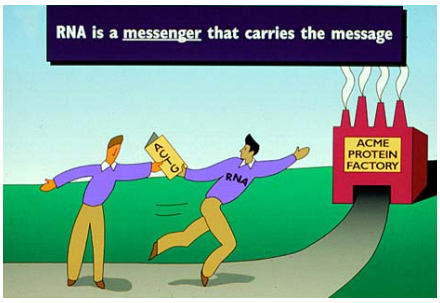

Figure 14
The RNA messenger then transfers the misspelled code, ACTG, to the protein factory in the cytoplasm

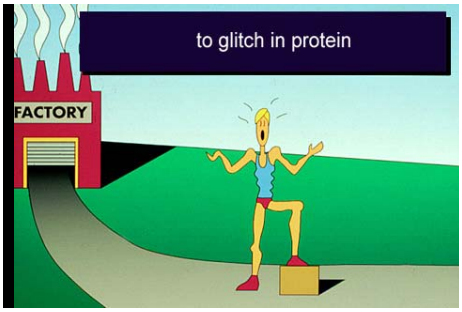

Figure 15
A glitch in protein

Now let's introduce another important concept in human genetics called linkage disequilibrium (Figure 16). This is a genetic term that has to do with the prediction that if two markers on a piece of DNA, let's say the yellow asterisk and any of the brown dots, are not linked, then there will be complete equilibrium in the genome between these markers. With equilibrium, you would find no examples where the brown dots would be linked in the same piece of DNA to the asterisks. But, on the other hand, if there is linkage disequilibrium, then all of these markers will be found together on the same piece of DNA. Lets say the brown dots and the asterisk represent single nucleotide polymorphisms, and that these five polymorphisms are close together in one piece of DNA in an ancestral chromosome. If the asterisk is associated with one or more of the brown spots in a present day chromosome, then these SNPs are considered to be in linkage disequilibrium. The presence and extent of linked markers in the human genome is an important discovery that has come out in the last few years – i.e., that the genome has blocks or units of DNA allowing markers such as SNPs to be in linkage disequilibrium.

Figure 17 shows us an example of actual results that can be obtained when we map single nucleotide polymorphisms in human beings. These results are from an important signaling gene called interleukin-6, and on the left you can see an example of what is found with the detection of six of the SNPs in this gene (the overall gene has 49 SNPs). Reading across the top, we have the identification numbers for the six SNPs. Reading vertically, we have the identification codes for a number of individuals – the matrix gives an actual representation of the genotyping measurements or data regarding their individual SNPs. What we can see right away in this so-called Visual Genotype is that the SNPs vary quite a bit in these individuals. The blue boxes mean that the individual has the common, "normal" base pair on both chromosomes – that's called homozygous. The red boxes mean the individual carries the normal base pair on one chromosome, but has a different base pair on the other chromosome – we call that heterozygous. As you can see, most of the boxes are blue, because most of us have the predominant or common base pair, while a few of us have a different base pair. There are some individuals who have the common base pair corresponding to all six SNPs, with six blue boxes reading across from left to right.

In some cases, the DNA information is missing, indicating that the analysis didn't work. (Technical challenges are a factor in genotyping science, also.) And, in some very rare cases, both of the chromosomes have a SNP. We don't see an example of this in the graphic showing the six SNPs. But, when we look over to the complete set of 49 SNPs, on the right-hand side, a couple of new points are readily apparent, including yellow boxes.

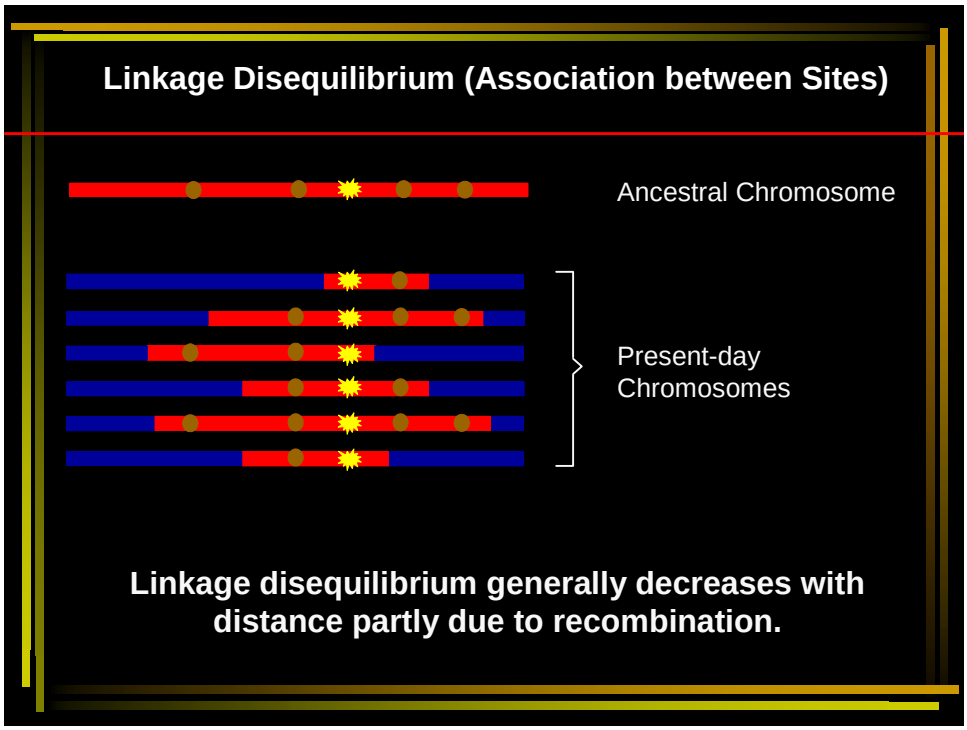

Figure 16
Linkage disequilibrium generally descreases with distance partly due to recombination

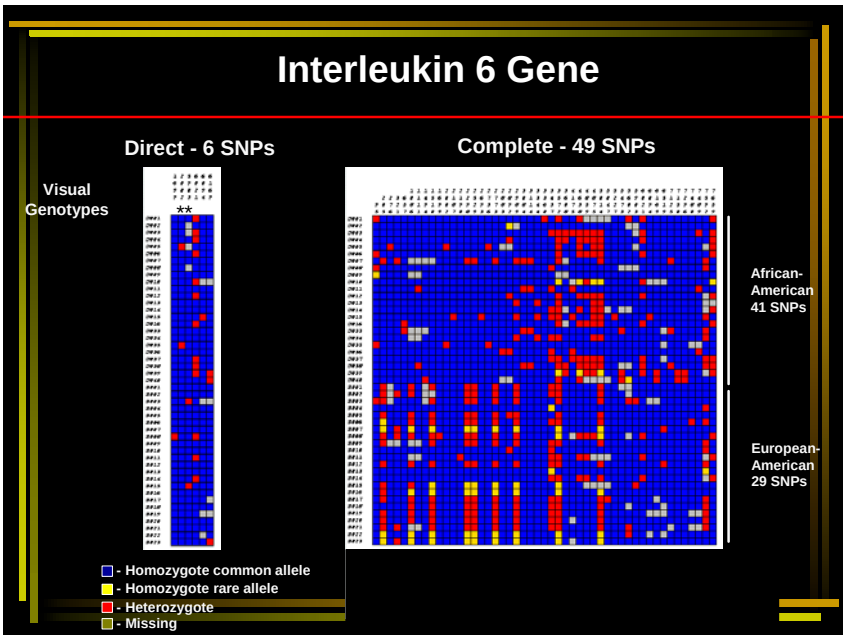

Figure 17
Interleukin 6 gene

First of all, as we look at the patterns for each individual, we can see that almost none of them are identical. This is to be expected – we're all different with regard to this type of analysis. Looking on the right-hand side of this panel, you can see that the investigators attempted to subgroup or stratify the identification of their study cohort into African-American individuals and European-American individuals. You can see that there are general differences between those two groups, for the individuals included in this particular study.

Now let's look in a little more detail with these data at this issue of linkage disequilibrium (Figure 18). Can any of the SNPs we see here on the left actually be placed in a linkage disequilibrium pattern, so that all of the SNPs segregate together? Yes, it does look like there are patterns here in the left-hand panel, and in the analysis shown in the right-hand panel, the investigators clumped the data together into patterns. In this analysis, you can see that there are some homozygous SNPs, shown in yellow, where both of the chromosomes have the base pair or SNP that is rare in the general population. The red boxes again show that many of the individuals have a SNP on one chromosome but not on the other. There are some examples where the data were not clear (grey). But, the other feature we see is that some of the individuals are similar. Once again, the classification or stratification as African-American and European-American is done on this slide. Obviously, there are some general pattern-level differences for these two groups in this study.

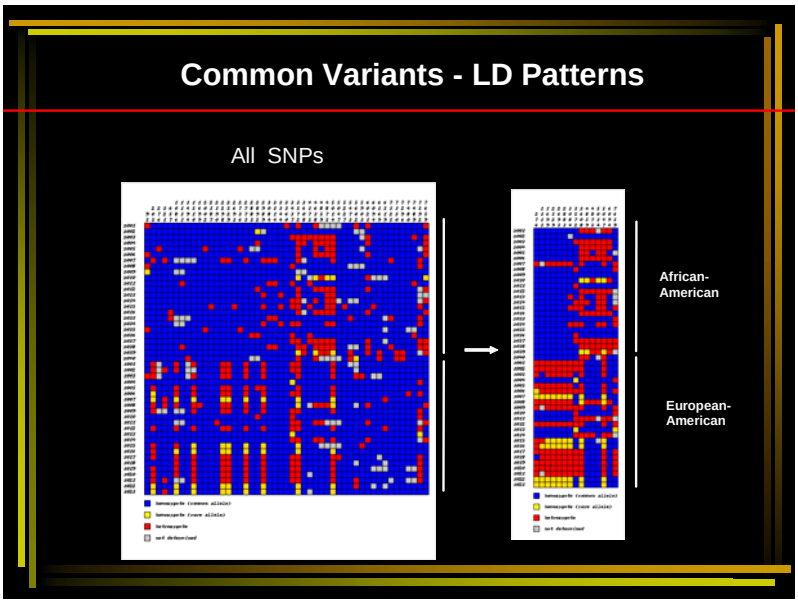

Figure 18
Common variants – LD patterns

Let's switch gears now and move into the area of environmental exposure and risk of disease. Now that we've learned about genetic factors, I'd like to introduce the concept of environmental exposures combining with genetic susceptibility to produce the bulk of risk for the common diseases.

Figure 19 illustrates that common diseases (the orange oval) result from the overlap of genetic and exposure to environmental factors; the greater the contribution to the health burden in the population due to common diseases the darker the color in the two larger ovals representing genetic factors (in red) and exposure (in blue). As shown environmental factors are broadly defined. The very strong genetic factors acting alone, account for a very small portion of disease in and of themselves as illustrated in light red on the left-hand aspect of the red oval in the graphic. So, for example, the strong genetic diseases like Huntington's disease represent a small portion of the overall public health burden. Most forms of cancer and the other common diseases would be in the overlapping central orange oval, because these diseases arise from a combination of genetic factors and environmental exposures.

In the case of environmental exposures, today, fortunately, most of the extremely toxic occupational exposures have been eliminated. For example, here in the U.S. we no longer have strong benzene exposures in the workplace or in our general environment. Hence, the amount of public health burden from acute toxicity due to exposure to environmental hazards alone is only modest. The major public health burden is in the middle, where exposure to environmental factors overlaps with genetic factors to explain common disease.

What do we mean by environmental factors? As you can see in Figure 19, in our view of the field of environmental health sciences we include standard items like toxicants, that is, man-made chemicals, and toxins, but we also include socioeconomic stress and other forms of stress, diet, medicines, lifestyle, and infectious disease history. So, the definition of environmental factors is broad.

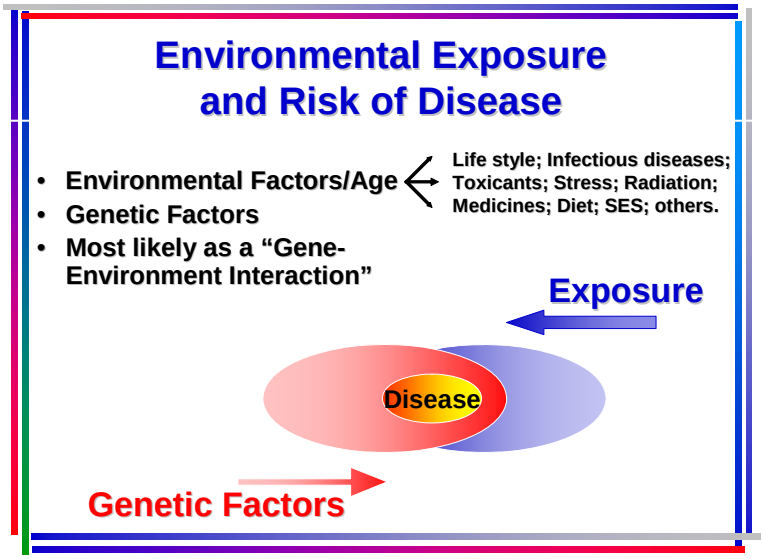

Figure 19
Environmental exposure and risk of disease

What this diagram really means is that the exposure situation alone cannot be tracked to the major public health burden from common diseases. And, the same point is true with the genetic factors or SNPs. But, the specific combination of an exposure plus a certain genetic variation leads to an increase in risk of disease. The effect, in many if not most cases, is specific to the kind of exposure and the kind of SNP.

It's easy to understand this latter point: Let's consider the DNA protective capacity known as DNA repair, where we know there are several different types of DNA repair pathways in our cells that don't overlap functionally. For example, when you're out in the sun and are exposed to UV light, you get a type of DNA damage in skin that is repaired by a certain repair pathway; if you have exposure to UV light and also have a SNP that inactivates this dedicated repair pathway, then this will lead to mutations and eventually to skin cancer. On the other hand, if you do not have a SNP in this repair pathway, the same sun exposure will be harmless. So, in this example, the combination of a specific type of stressful exposure and a deficiency in the specific cellular pathway for protecting against that exposure are key; this type of combination situation is the essence of the gene-by-environment or gene-environment interaction concept in understanding causes of common human diseases.

There are already many examples in the literature from molecular epidemiology studies where a gene-environment interaction is necessary to observe a higher adverse risk. Just one example is with several SNPs in the XRCC1 DNA repair gene combined with cigarette smoking; this combination has been associated with an increased or decreased risk of cancer, depending on the SNP.

What are the future directions in this field of gene-environment interactions? The science of using alternate experimental models, especially animal models and mammalian cell models containing or representing the human SNPs is a huge opportunity. We will also see more comparative genomics, i.e., use of alternate experimental systems, where we can identify what the functional significance of a SNP is, by virtue of studying other systems and comparing them to the human situation. Structural biology will also expand our information, where we will

analyze atomic level structures to learn what's wrong with a protein that contains a SNP. Of course, in vitro studies such as enzyme assays and the reconstitution of biological pathways will also be important. All of these areas will yield useful new knowledge, but the topic I want to focus on now is one you're going to hear a lot about in the future: Whole genome association, which is becoming the cornerstone of disease susceptibility research.

In the current issue of Nature Genetics, you can find papers on whole genome association, where SNPs spanning the entire genome have been probed. In such studies, one is able to ask the question, "Is there any correlation between a disease and a SNP or SNP pattern?" The answer, of course, is" yes," such correlations are being found. For example, it's was reported that there are five SNPs associated with a form macular degeneration. But, if all five of them are not considered as a unit in the analysis, the association broke down. Thus, this common condition appeared toappeared to be a multigenic effect discovered through whole genome association research. This type of information is developed in the scientific field called molecular epidemiology. It is a field that makes use of a combination of genetics and genomics research with epidemiology. This approach is how we're going to make big advances in the future and get wonderful surprises, as to genetic associations with disease as a function of environmental exposures.

So far in this talk, we've seen that we're in a new age of genomics, where there are many new technologies such as whole genome association and new environmental technologies, as well, to give us better precision, to open up windows that we never before could see through. As I said above, there are many examples now where combinations of a genetic variation (SNP) and exposure to an environmental factor have been found to impact human health. We are, indeed, in a new gene by environment age, and this will pose new challenges for all of us.

Again, where is all of this going? Obviously, we will eventually be able to individualize disease risk assignments based on our SNPs and our environmental exposures, and use that information for prevention, diagnosis, and therapy. This information is going to enhance the efficiency of therapeutics and of safety assessment for preventing exposure and disease. It's going to enhance the efficiency with which we can do toxicity assessment. For example, if you want to know whether your water contains arsenic, this kind of hazard assessment can be vastly improved with the new –omics technologies. But the point I really want focus on now is that we need better policies for protection of individual privacy and dignity, so that we can actually make full use of this new the omics-based opportunities to protect the environment, improve human health, and better prevent and treat disease.

What is the timeline on use of omics technologies to improve human health? One can hear a lot these days about individualized medicine and so on, and individualized risk assessment. But actually, even with the explosion in the omics sciences, it will be a long time before we can routinely reduce the new science to practice in the doctor's offices and hospitals and clinics. Yet, better efficiency in drug design and in drug safety assessment will probably be with us within the next several years. Better efficiency in toxicity assessment is also feasible and is probably five years away. What about better policies? We need these right away, right now, in order to protect the science and to protect society.

We face many new challenges with this new omics science. For example: How do we better understand ways to apply the new science to improve health for everyone? What approaches can we take toward educating ourselves about the pros and cons of genomics? What new policies do we need? And, how do we develop the information and policies without offending stakeholders? How do we provide equal and appropriate access? As this new science emerges, these topics must be addressed, because the science will not be allowed to realize its full potential to improve human health if we don't address them.

Let's look at just one example. I've talked about the Human Genome Project – a tremendous success in biomedical research applying around the world, and in addition there is the Haplotype Map program. This involves a mapping of SNPs across the entire genome. This advance means that in the near future there could be a form or version of your SNP genotype in a database. And mine too. All of us may have our SNPs filed in databases in the near future. We're going to be able to examine disease associations in individuals by making use of these SNP patterns or whole genome associations, and we are going to discover all sorts of new things that we haven't appreciated in the past, as to genetic and environmental factors in disease risk. We are eventually going to be able to do this with measurement of specific SNPs that will allow routine screening in clinical laboratories and commercial laboratories across the country and around the world.

As I see it, there are two big challenges I'd like to call your attention to. One: who will be left out in the application of this revolutionary analysis? If we leave certain people out, then it's just not going to work; it's not going to be appropriate. There's a need for health disparities and other awareness, to make sure we can apply this new science appropriately and evenly. The second challenge I am concerned about is DNA-based "labeling" of individuals and groups. It is a common practice in the research community to stratify or lump individuals into racial groups, and then at the same time to associate that information with DNA analyses. I have concerns about this approach, because I think it could lead to just another form of labeling. We're all individually different in our SNPs and in our genotypes and in our exposures, and there's reason to allow SNP analysis to become transformed into a way of harmful labeling.

How can we better understand and apply this new science and move beyond these concerns I just mentioned? What are some of the other concerns? For example, what are the approaches we need to take – and I think it's up to us to do this – to educate the general community about the new science? What new public policies, laws and guidelines do we need? How do we avoid doing harm? At the same time, how do we promote widespread and robust use of these new tools?

There are many policy issues to be dealt with in the near future as the new science emerges. First of all, we need good solid ways of obtaining informed consent. Obviously, in doing the whole genome associations that I've discussed, scientists need to have informed consent in order to look at your DNA and mine. Thus, when the information goes into a database, there needs to be an informed consent process as to the privacy and use of that information. Also, we must have community involvement in understanding and prioritizing the research. We need to work on this a lot. We need to fill the gaps in communication of information to the public. Many people in the general public are not aware of much of the information I've discussed in this presentation, and we need to learn how to communicate these concepts much more effectively. Conflict of interest is another important issue. For example, what if doctors working on this topic own stock in companies that market genomics products and recommend the products to their patients? This is just one example among many of potential conflict of interest problems. We must find ways to avoid these problems. In both our research and our communications efforts, we need to maintain an awareness of cultural issues. And we need to be diligent in ensuring that people's expectations are realistic and appropriate regarding outcome and follow up for this new type of gene-environment information. Again, the key factor is communication. Finally, we need a formal process for policy development to protect and enable this research.

These issues regarding policy raise a variety of concerns. Again, perhaps the most important is the protection of privacy and confidentiality within the genetic and genomic databases and with truly informed consent. That will be crucial to protect us all against adverse use of the information. For example, once our susceptibilities are known, what will prevent discrimination in the workplace, by employers or insurance companies or health care providers? Another concern is that we will require

protection against "transfer of responsibility" because of knowledge about SNPs. What do I mean by this? At the moment, it is the responsibility of the city, state, or nation to protect us from certain types of environmental hazards. But in the future there may be attempts to shift that responsibility to individuals who are known to be unusually susceptible to a certain type of environmental stress. Or the environmental regulation standards could be changed in such a way that highly susceptible individuals would not be protected from specific hazards. As a society, how far do we go in making such protection of susceptible individuals a civic responsibility?

As we develop new policies addressing these issues and concerns, "how much is enough?" This is the standard, old-fashioned question, and the question about the role of the precautionary principle that we deal with in the environmental field. How do we know when we have enough information in order to act to create a policy? How do we achieve agreement so that public officials can successfully develop and implement policies? And, do we have mechanisms in place to meet these policy needs? I believe this area of "policy science" surrounding gene-environment interaction research is going to be one of the most important fields in the next ten to twenty years, and probably beyond. For those of you interested in public policy, especially policy as it relates to human health and disease, there is a lot of work to do in the area, with a lot of very interesting and important questions to address.

Acknowledgements

The views expressed in this article are the author's and his alone and do not represent the official position of the NIEHS.

I wish to thank Debbie Nickerson, Bill Suk, and Francis Collins for slides, and I thank Bill Suk, Kenneth Olden, Debbie Nickerson, Lovell Jones and Joseph Lowery for helpful discussions. I also thank Ernie Hood for excellent editorial assistance and Sally Tinkle for reading the manuscript.

Author Information

Samuel H. Wilson
National Institute of Environmental Health Sciences
National Institutes of Health
P.O. Box 12233
Research Triangle Park, NC 27709

* corresponding author