

STORIES OF DISCOVERY

Single Nucleotide Polymorphisms (SNPs) and Haplotypes: New Tools for Tracing Inherited Diseases

Since the beginning of self-awareness, humankind has been interested in our heritable differences. Many are of no consequence, but a few play a role in risk of illness – put another way, about all illnesses tend to run in families. The research disciplines of population genetics and genetic epidemiology have been studying these phenomena for decades; but now, courtesy of the Human Genome Project, they are rapidly acquiring a molecular foundation.

Any two people have the same DNA sequence for about 99.9 percent of their DNA. The 0.1 percent difference includes genetic variation that can lead to differences in the risk of getting various diseases and having an adverse drug reaction. Some diseases, such as cystic fibrosis and Huntington's disease, result from differences in the DNA sequence in single genes. However, many common diseases such as diabetes, cancer, heart disease, psychiatric disorders, and asthma are influenced by complex interactions between multiple genes as well as by non-genetic factors such as diet, exercise, smoking, and exposure to toxins. With the tools of the Human Genome Project, finding the genes for diseases caused by alterations in single genes has become relatively straightforward. Finding the genes that contribute to common diseases remains extremely difficult. To make this task easier, faster, and more efficient, the Human Genome Project is creating a new set of powerful genomic tools that, together with the reference genome sequence, will dramatically enhance scientists' ability to identify the genetic contributors to a host of common diseases.

The Human Genome Project is creating a catalogue of the places in the genome where the DNA sequence differs among individuals. The most common variations are single nucleotide polymorphisms (SNPs), or places where the DNA sequence varies by a single base pair or DNA letter. Between two unrelated individuals, these occur approximately once every 1000 bases. The availability of the working draft of the human genome sequence has dramatically accelerated the identification of SNPs in the last year.

The rapid pace of SNP identification has been facilitated by the NIH DNA Polymorphism Discovery Resource (PDR). The PDR provides a common set of 450 DNA samples to scientists seeking SNPs. The samples were collected under strict ethical guidelines from anonymous unrelated U.S. residents of diverse ethnic backgrounds (http://www.nhgri.nih.gov:80/About_NHGRI/Der/variati.htm).

NIH has funded studies to identify SNPs, and this effort has been complemented by the SNP Consortium (TSC), a non-profit consortium whose members include the Wellcome Trust, 10 pharmaceutical companies, IBM, Motorola, and Amersham. The goal of the consortium is to develop a high-quality SNP map of the human genome and to make the information related to these SNPs available to the public without intellectual property restrictions (<http://snp.cshl.org/>).

Through these combined efforts, nearly 3 million SNPs have been identified and are available in public databases for use by scientists in industry and academia.

The next step of the Human Genome Project is the generation of a new map of the patterns of genetic variations across populations, a haplotype map. The variants do not occur at random, but are correlated in important ways with the neighbors. By knowing the pattern of variation along chromosomes, scientists can select a modest number of SNPs distributed along the chromosomes that will reveal the underlying pattern of variation. Thus, testing for a single SNP will reveal the information for a large region surrounding the indicator SNP.

This pattern of genetic variation and set of indicator SNPs will permit scientists to scan the entire genome to find chromosomal regions that are statistically associated with disease. Scientists can then refer to the working draft of the human genome to find the specific genes in that region and narrow their search for the disease causing alteration. Whole genome association studies such as this have heretofore been too expensive. The availability of haplotype maps will make these studies faster, more efficient, and much more affordable.

In July 2001, the NIH held a meeting that brought together the leading experts in the world to discuss how haplotype maps would be useful for mapping genes contributing to disease, the methods for constructing such maps, the data about haplotype structure in populations, the types of populations and samples that might be considered, the ethical issues, and how such a project should be organized.

While still in its formative stages, the planned development of a haplotype map will provide a critical resource for researchers seeking to identify and understand the genetic basis of common human disease.

Mass Spectrometers Weigh the Evidence for Health and Disease

Preoccupation with weight is widespread in America. Everyone knows that the body's mass affects the way one feels, functions, and interacts with others, and physicians use weight as a quick indicator of an individual's health, activities, and endurance. What is less well known is that mass is similarly informative at the molecular level. Scientists have long recognized that atoms and molecules have distinctive weights, but only recently have they discovered that molecular mass also offers insight into the functions, interactions, composition, and overall shape of complex biological molecules like proteins and DNA. These discoveries, enabled by technological advances in molecular "weighing" machines known as mass spectrometers, have contributed to improved diagnostics, treatment, and understanding of diseases such as cancer, AIDS, and prion-related disorders like mad cow disease.

Over the years, mass spectrometry has proved to be a remarkably versatile tool, enabling an enormous assortment of scientific advances in astronomy, physics, environmental science, and medicine. The technique is used by hospital staffs to identify unknown toxic compounds in a patient's blood or urine, and at the Olympics to detect illegal use of steroids by athletes. Today's state-of-the-art mass spectrometers can identify and characterize hundreds of complex proteins at near-instantaneous speeds, making mass spectrometry a star player in the emerging new discipline known as proteomics. The goal of proteomics is to detect and evaluate the thousands of proteins expressed by the genome and then understand how all these molecules work in concert to maintain life.

The first mass spectrometer was created nearly a century ago by distinguished physicist and discoverer of the electron, J.J. Thomson. Although his primitive instrument bears little resemblance to today's sleek and sophisticated models, the underlying principles are remarkably similar and simple. Then as now, sample molecules in a mass spectrometer are first vaporized and then ionized, or broken down into smaller bits that have either a positive or a negative charge. The ions are then subjected to an electromagnetic field, which can have varying effects on their speed, trajectory, or energy, depending on each ion's mass and charge.

The mass spectrum itself is akin to the colorful spectrum produced when sunlight passes through a prism. But whereas a prism separates light into a continuum of distinctive colors of varying wavelengths, a mass spectrometer splits molecules into distinctive ions and focuses them along a continuum of varying masses. Because mass spectrometers also measure the relative abundance of each ion, scientists can determine the overall mass of a molecule, the mass of each ion piece, and the relative abundance of each piece, all of which provide clues to the composition and structure of compounds under study.

Originally the province of physicists, mass spectrometers in the first half of the century were massive, arcane instruments capable of analyzing small carbon-based compounds or simple organic mixtures like petroleum. But for more than four decades, NIH-supported scientists have incrementally enhanced the sensitivity and flexibility of mass spectrometry to address some of the most pressing problems related to human health. Their progressing work has transformed

mass spectrometry from a technique accessible only to experts in engineering and gas-phase chemistry to a routinely used tool indispensable to today's biomedical research.

Some of the first breakthroughs on the biological front occurred in the 1960s at the Massachusetts Institute of Technology (MIT), where Klaus Biemann and his colleagues paired mass spectrometry (MS) with gas chromatography (GC), a technique that separates different types of molecules in a mixture. With NIH support, the MIT lab pioneered the clinical application of GC/MS by providing Boston-area hospitals with analyses of blood and urine from unconscious patients and children with suspected metabolic disorders. More sensitive than other techniques, GC/MS could rapidly identify even small amounts of toxic substances or abnormal metabolites in body fluids, allowing physicians to select appropriate, sometimes life-saving, therapies. Demand for these analyses expanded so rapidly in the 1970s that commercial laboratories eventually began to offer GC/MS services to hospitals nationwide.

Biemann's laboratory also devised innovative computer systems for acquiring, storing, and processing GC/MS spectra. The resulting searchable databases could quickly identify unknown compounds by comparing them to known spectra. Today, expanded versions of these digital libraries are a mainstay of clinical diagnostics and research.

Mass spectrometry took another leap ahead in the 1980s when improved techniques for ionizing molecules revolutionized the study of large, fragile compounds like proteins. Earlier ionization protocols were so harsh that proteins were often destroyed before they could be effectively ionized. Proteins amenable to ionization with the old techniques often splintered into fragments that were either too small and numerous for successful analysis or too large and unmanageable to focus into a detectable spectrum.

But the so-called "soft" ionization techniques known as matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) were gentle enough to systematically dismantle large proteins and sensitive enough to study their structures at high resolution. While previous MS protocols were limited to analyzing proteins with masses up to about 20,000 daltons, MALDI and ESI could interrogate proteins with masses up to several hundred thousand daltons, and could do so accurately even with vanishingly small samples.

These and other advances have opened a new world of opportunities for biomedical scientists. At the University of California, San Francisco, mass spectrometry uncovered the detailed three-dimensional structure of the prion protein, responsible for mad cow disease and devastating human brain disorders. Scientists at the University of Georgia used mass spectrometry to examine complex sugar-studded proteins on the surface of the human AIDS virus, suggesting that these molecules help to camouflage HIV and evade the body's immune system. And several teams of scientists have employed mass spectrometry to compare very small quantities of proteins in both tumor and normal cells, with the ultimate goal of identifying potential drug targets.

By coupling mass spectrometry with today's powerful computers, NIH-supported investigators have also developed high-throughput techniques that are a cornerstone in today's proteomics

investigations. These sophisticated systems can obtain mass spectra at record rates exceeding 100 acquisitions per second, and then make intelligent decisions about which ions in a spectrum should be selected for further analysis. Using these integrated technologies, scientists have systematically analyzed proteins expressed in various cells or tissues, and have identified multiple protein components of critical cellular structures like the ribosome or nuclear pore complex. These detailed investigations offer exciting possibilities for development of highly targeted, highly effective new therapies.

From its humble beginnings 100 years ago, mass spectrometry has matured into a sophisticated, sensitive, and indispensable tool with surprising versatility. It raises the threshold for what can be discovered and accomplished in biomedical science, and it contributes to a broad and detailed picture of the complex underpinnings of life. The landmark discoveries achieved with J.J. Thomson's primitive instrument could not have even been imagined in his lifetime. Scientists today expect that mass spectrometry will continue to grow and expand in unforeseen ways, and offer significant opportunities for enhancing human health.