

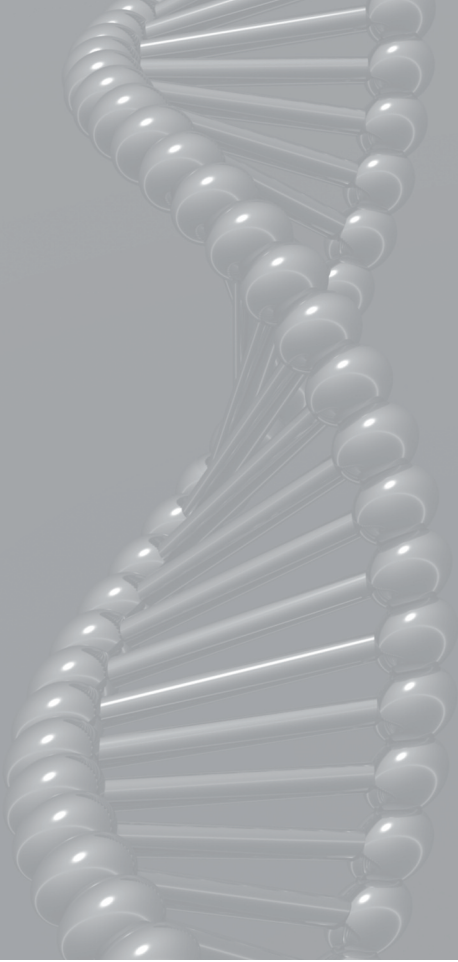


## Genome-Wide Association Studies and Single Nucleotide Polymorphisms

**Research Advocacy Network**

*Advancing Patient-Focused Research*

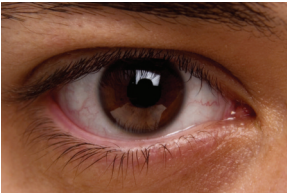
[www.researchadvocacy.org](http://www.researchadvocacy.org)



# Genome-Wide Association Studies and Single Nucleotide Polymorphisms (SNPs)

## Introduction

The human genome, or all of our DNA, contains the genetic blueprint for the human species and the genetic information we pass from one generation to the next. Interspersed within long stretches of DNA are about 20,000 to 25,000 human genes. The exact structure of these genes varies between people, explaining differences in eye color, height, and predisposition to certain diseases like cancer.



As you might imagine, researchers are keenly interested in identifying the genetic differences that predispose some of us to cancer and others to Alzheimer disease, heart disease, or diabetes. In the past, studies could evaluate only one or a few genes at a time. Researchers had to guess which genes to analyze based on existing scientific results. Although this type of research led to breakthrough treatments for a number of life-threatening diseases, the strategy has not been as fruitful for complex diseases such as cancer that often involve multiple genes.



This is where genome-wide association studies—sometimes called G-WAS (gee-wahs)—have proven so useful. These studies allow researchers to study differences in human genes by searching thousands or even millions of differences in DNA at the same time. As we will discuss in the following text, genome-wide association studies do not look at entire genes in each person. Instead, they look for variations in the DNA referred to as single nucleotide polymorphisms, or SNPs (pronounced “snips”).

In the following text, we will explore SNPs and genome wide association studies in more detail. We will first review the basics of DNA as a foundation to understanding SNPs. We will then discuss genome-wide association studies, including relevant definitions, study procedures, and types of information that can be obtained. Finally, we will consider how genome-wide association studies may benefit us in the future as part of the trend toward personalized medicine and some real-life examples of genome-wide association studies in breast cancer.

## The ABCs of DNA

We’ve all heard of the letters DNA and most of us have at least a vague understanding that DNA is somehow related to our genes. DNA is the abbreviation for a chemical called deoxyribonucleic acid and it is, in fact, the material that makes up our genes.

DNA consists of a series of 4 chemical bases that occur in a specific order and are paired with one another. The 4 bases, known as nucleotides, are adenine, thymine, guanine, and cytosine, but they are usually referred to by the first letter of their names: A, T, G, and C. Each of these nucleotide bases pairs with another base in a specific manner: A only pairs with T and G only pairs with C.

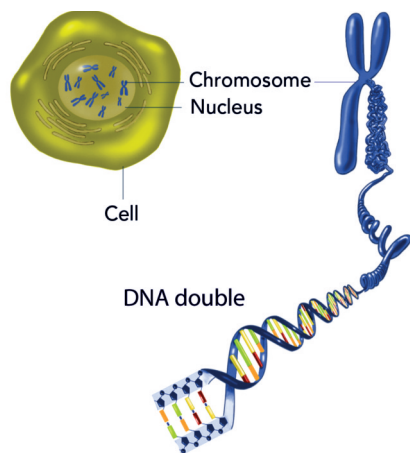
### Nucleotide bases

The chemicals that make up DNA. Their names are adenine, thymine, guanine, and cytosine and are most commonly referred to by the letters A, T, G, and C.

### Nucleotide base pairs

Along our DNA, each nucleotide base pairs with only one other base: A pairs with T and G pairs with C. These are called base pairs and there are 3 billion of them in the human genome.





Inside cells, DNA typically exists in two connecting strands called a double helix. The coiled strands of DNA can be found in nearly every cell in our bodies contained within structures known as chromosomes. For more information on chromosomes and the double helix structure of DNA, you may want to visit the National Human Genome Research Institute Web site at <http://www.genome.gov/10000202> or the National Cancer Institute Web site called Understanding Cancer at <http://www.cancer.gov/cancertopics/understandingcancer/geneticvariation/AllPages>.

## DNA Variation

Each one of us is 99.9% genetically similar to every other person on the planet. The differences in our DNA, which amount to 0.1%, are referred to as DNA variation. Many types of variations can occur, most of which do not lead to any noticeable differences between people. However, some of the variations can actually be beneficial, whereas others may increase our risk for disease. Variations can be caused by a change, addition, or removal of one or more nucleotide bases. Other types of variations affect parts of chromosomes or even entire chromosomes.

### Example of a SNP

Hypothetical DNA sequence in a gene of most individuals:  
**TCCGAAGA**

Hypothetical DNA sequence in the gene showing a SNP:  
**TCTGAAGA**

### Polymorphism:

A change in the nucleotide base sequence of our DNA that occurs in at least 1% of people.

### Single Nucleotide Polymorphism (SNP):

A polymorphism that occurs in one single nucleotide base pair.

### *Single Nucleotide Polymorphisms (SNPs)*

Here we will focus on a type of DNA variation known as a polymorphism that is relatively common in the population. A polymorphism is a change in the nucleotide base sequence of our DNA that occurs in at least 1% of people. Different types of polymorphisms can occur in our DNA, including the deletion of base pairs, the insertion of extra base pairs, or the exchange of one nucleotide base for another. When a polymorphism involves the exchange of one single nucleotide base pair, it is called a single nucleotide polymorphism or SNP. SNPs are the most common type of genetic variation, accounting for 90% of the differences in the human genome.

SNPs may or may not have any noticeable consequences to our health. We will focus here on the SNPs that do have an impact on our health because it is these SNPs in which scientists, patients, and advocates are interested.

SNPs may be associated with our health in several ways. Some SNPs may be associated with a predisposition to disease, as is often the case for cancers. SNPs may also be associated with how fast a disease progresses, the response of a disease to a given treatment, and drug side effects.

### *An Example of SNPs that Predict Disease Risk*

Alzheimer disease is an example of a condition in which SNPs have been identified that affect disease risk. Different forms of the gene called apolipoprotein E (ApoE) increase or decrease a person's risk of developing the disease. The ApoE gene contains SNPs at two locations that result in three possible forms of the gene, which are referred to as E2, E3, and E4. People with at least one copy of the E4 form of the gene are at increased risk of Alzheimer disease, whereas people with at least one copy of the E2 gene are at decreased risk.

Not all people who have copies of the E4 form of the ApoE gene will develop Alzheimer disease and not all people who have copies of the E2 form will be protected. Alzheimer disease, like cancer, diabetes, heart disease, and other common but complex human diseases are typically caused by multiple genes interacting with environmental factors. Diet, exercise, cigarette smoking, and ultraviolet light are all examples of environmental factors that may interact with genes to determine whether someone will develop cancer.

### How Is a SNP Different From a Mutation?

We have already defined a SNP as a change in a single nucleotide base pair that occurs in more than 1% of the population. By population-based standards, each SNP is relatively common, affecting more than 1 in every 100 people. Mutations are another type of genetic variation. They differ from SNPs in that they are less common. **Mutations are changes in the nucleotide base sequence of our DNA that occur in less than 1% of the population.** The term mutation is generally used to refer to a change in DNA sequence that has deleterious consequences for the organism, whereas a SNP can have either positive consequences, negative consequences, or no consequences.

We should note here that there is not universal agreement as to what the term mutation means. Many different definitions exist. For instance, some sources indicate that any genetic change having potentially negative consequences for the individual is a mutation. Other sources use the terms mutation and SNP interchangeably. Scientific sources agree, though, that if a genetic variation occurs in 1% or more of the population, it is a SNP, whereas if it occurs in less than 1% of the population, it is a mutation. The University of Utah's Genetic Learning Center Web site provides a more detailed explanation of the differences between SNPs and mutations: <http://learn.genetics.utah.edu/content/health/pharma/snips/>.

Variations in the breast cancer-related genes known as BRCA1 and BRCA2 are examples of mutations because they affect less than 1% of the population. Individuals with certain mutations in these genes are more likely than the rest of the population to get breast cancer, ovarian cancer, or certain other cancers. However, as with many cancer-related genes, mutations in these genes do not guarantee that an individual will develop these cancers and lack of mutation does not guarantee of protection against the disease. This is also true for SNPs.

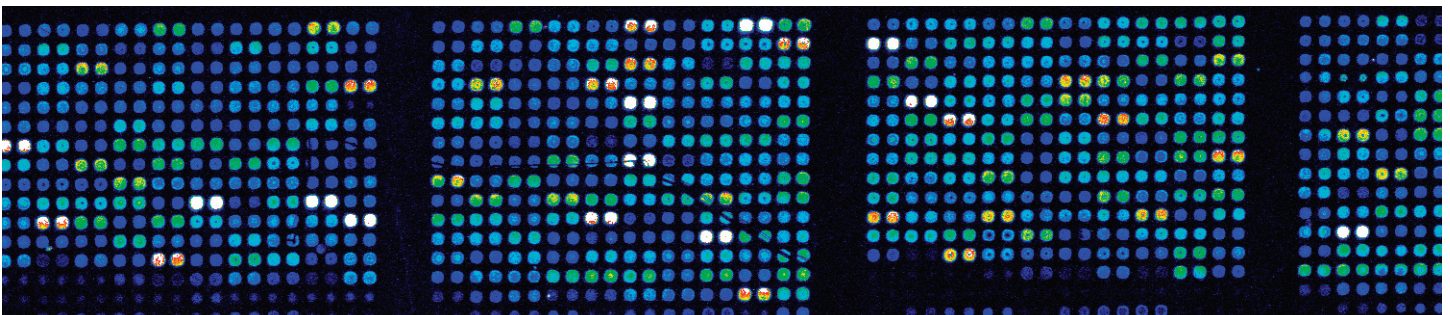
### Genome-Wide Association Studies (GWAS)

Genome-wide association studies involve scanning the entire genome of many people to look for SNPs associated with disease or its treatment. In this section, we first describe the general procedures used to conduct genome-wide association studies. We then consider the different types of information that can be obtained from these studies and provide some examples. You might hear these studies referred to as GWAS.

#### *General Procedures*

In a typical genome-wide association study, SNPs are compared between two groups of people that differ in some way, such as those with cancer and those without. The study begins with each individual providing a DNA sample, usually obtained via a blood test or cheek swab. These samples contain cells from which DNA can be extracted. Once the DNA is obtained from the cells and is purified, it is placed on tiny chips that use the chemical pairing between nucleotide

DNA Microarray – A  
Common Genomics Tool



bases to determine whether certain SNPs are present in the DNA. Researchers must decide which SNPs to search for based on available scientific information. After allowing time for the nucleotide bases to pair, the array is then processed so that only the paired bases will “stick.” This results in a colored circle on the array. The array is then scanned and analyzed by a computer that determines which SNPs are present.

Scientists then compare results for the two groups to determine whether any SNPs are substantially higher or lower in one group than the other. Some differences between groups will occur just by chance, just as we may occasionally meet another person who has the same birthday as we do. To minimize these chance coincidences, statistics are used to determine whether the differences between groups are large enough to be considered “real” effects. The SNPs that meet these statistical criteria are then considered possibly associated with the disease. To be more certain about the association between particular SNPs and disease, researchers must attempt to replicate the findings in a different population of patients.

Once a SNP of interest has been identified, scientists must try to find the gene with which it is associated. Databases containing the sequence (ie, the order of As, Ts, Cs, and Gs) of the entire human genome are available.

When SNPs have been definitively associated with a variable of interest, such as disease risk, they can be used to guide research and, eventually, health-related decisions. In the future, analyses of our SNP patterns will probably become a routine part of our healthcare, much as cholesterol levels and blood pressure are today.

## Types of Information Obtained From Genome-Wide Association Studies

### *Disease Risk*

One of the most common reasons for conducting a genome wide association study is to identify risk factors for disease. In this type of study, a group of individuals with the disease is compared to a group without the disease to identify significantly different SNPs.

We have already seen that SNPs in the ApoE gene can be used to help predict a person’s risk for Alzheimer disease. As an example in cancer, nearly 40 SNPs have been associated with the risk for prostate cancer. Some of these SNPs are better predictors of risk than others, and there is a trend now toward identifying a panel of SNPs, or multiple SNPs, that are better than any single SNP at predicting disease risk.

### *Prognosis*

Another potential use of SNPs is in disease prognosis, or the predicted course of disease in the absence of treatment. Some people have cancers that are more aggressive than others, even if the cancer affects the same organ such as pancreas, liver, or brain. The aggressive cancers grow more quickly and are more likely to spread to other parts of the body.

In some cases, SNPs can help predict whether a cancer is slow-growing or aggressive. This type of knowledge can help determine which treatment a person selects. An example of a SNP that may be useful for predicting the severity of disease is in the CYP3A4 gene. This gene encodes a liver enzyme that helps break down testosterone. Certain SNPs in the CYP3A4 gene have been associated with increased severity of prostate cancer.

**Prognosis:**  
the predicted course  
of disease in the  
absence of treatment.

**Pharmacogenomics:**  
the study of how a  
person's genes  
influence his or her  
responses to drugs.

### *Pharmacogenomics*

Pharmacogenomics is the study of how a person's genes influence his or her responses to drugs. Scientists have found that the pattern of SNPs can affect how someone breaks down or metabolizes drugs, which can influence the intensity of side effects a person experiences. Another research focus in pharmacogenomics is the effect of SNPs on treatment response, or whether or not a person will respond to certain drugs.

An example of how genetic variation can influence drug metabolism involves the enzyme thiopurine S-methyltransferase or TPMT. Some mutations in this enzyme reduce a person's ability to metabolize drugs known as thiopurines, which are used for the treatment of leukemia. People with TPMT mutations could die if given normal doses of thiopurines because the drugs become too highly concentrated in their bodies. In contrast, people with these mutations can usually safely tolerate lower doses of thiopurines.

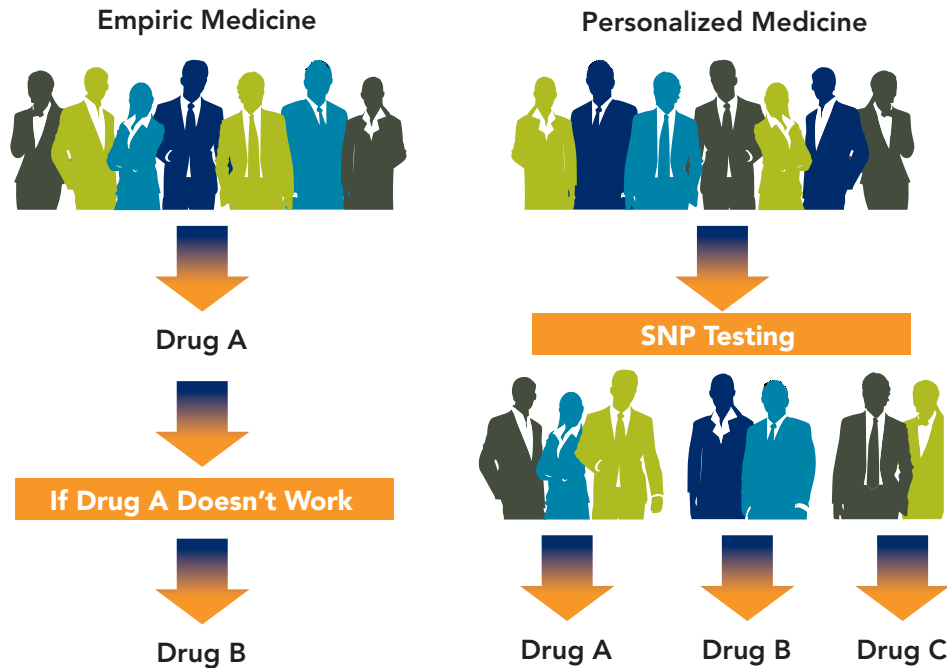
Another example of a genetic variation that can influence drug metabolism involves a chemotherapy drug called irinotecan that is used to treat some types of colorectal cancer and other cancers. Individuals with certain mutations in an enzyme known as UGT1A cannot break down irinotecan as efficiently as others, and may develop toxicities or side effects such as diarrhea as levels of the drug accumulate in the body.

Scientists are also trying to determine which SNPs can help predict whether someone will respond to or benefit from a cancer drug. Research is underway in this area, and preliminary findings have found associations between certain SNPs and drugs responses. However, this research is still in the early stages. Subsequent studies are needed to validate the findings before the SNPs can be used in clinical cancer treatment.

### *Personalized Medicine*

An important prospect for the future is the use of genomic technologies such as SNP profiles to help individualize or personalize our medical care. As technologies become more comprehensive and less expensive, SNP testing will likely become a routine procedure at the doctor's office. If SNP tests can tell us about our disease predispositions, we may be able to take steps to avoid certain conditions; for instance, a person prone to lung cancer would have extra incentive not to smoke. SNP tests will also help physicians characterize our disease more precisely. Knowing the specific genetic features of each person's cancer will allow physicians to better match the treatment to the individual's profile, perhaps increasing the effectiveness of therapy and helping minimize serious side effects.

Personalized medicine contrasts with the type of medicine that has been practiced in the past, sometimes referred to as empiric medicine. In empiric medicine, each person with the disease is treated with the drug that works best for most people. If that doesn't work, the person then receives a different drug to see if that works, and so forth. With personalized medicine, people are given SNP tests first and then a drug is selected to match their specific disease profile.



### Examples of Genome-Wide Association Studies in Breast Cancer Research

To see how genome-wide association studies are being used in breast cancer research today, we now turn to several real-world examples. This research involves a drug that blocks a key chemical the body uses to generate new blood vessels. This chemical is known as vascular endothelial growth factor or VEGF (vej-eff). Tumors need new blood vessels to sustain their growth, and certain drugs that interfere with this process have shown effectiveness for several different types of cancers. One of these drugs is bevacizumab (Avastin®). Bevacizumab binds to VEGF, interfering with its activity.

Although bevacizumab works for some cancers, its results in breast cancer have been equivocal: In some studies it seems to be effective and in other studies it does not. Additionally, the drug causes serious side effects in some patients. Some researchers have suggested that, although the drug is not beneficial for all breast cancer patients, there may be a subgroup of patients that does benefit. Moreover, there may be subgroups of patients that are more likely to experience serious side effects, whereas others may tolerate the drug fairly well. Researchers have turned to genome-wide association studies to help define the various subgroups of patients.

Dr. Bryan Schneider and his colleagues from Indiana University and other cancer research centers are conducting genome-wide association studies on DNA samples from patients who have received chemotherapy for human epidermal growth factor receptor 2 (HER2)-negative breast cancer as part of clinical trials. One of these studies found that patients with a certain SNP profile in genes related to VEGF showed significantly better overall survival than those with a different SNP profile when bevacizumab was added to their chemotherapy regimen. The researchers also found that two different SNP profiles were associated with reduced tendency to develop high blood pressure when bevacizumab was added to the chemotherapy regimen. These results suggest that breast cancer patients with certain SNP patterns may respond better to bevacizumab, whereas others may be less likely to develop high blood pressure.

In another study, researchers are examining whether the addition of bevacizumab to the current standard chemotherapy regimen (doxorubicin, cyclophosphamide and paclitaxel) reduces the risk of cancer recurrence in women who have undergone surgery for HER2-negative breast cancer. Dr. Schneider and his colleagues have found that patients with one of several different SNP patterns are at increased risk for a side effect known as neuropathy or nerve damage. Neuropathy may be caused by the chemotherapy drugs paclitaxel and docetaxel and may be manifested as pain, decreased reflexes, and abnormal sensations of the hands and feet. If further research confirms these findings, it may eventually be possible to help optimize treatment for patients with the two SNP profiles by altering their chemotherapy regimen.

### Summary and Conclusions

Genome-wide association studies are an increasingly important part of biomedical research. These studies evaluate SNPs in our DNA that may be associated with some scientific or medical variable of interest, such as disease predisposition, prognosis, or response to drugs.

SNPs or single nucleotide polymorphisms are variations in the DNA that occur at a single nucleotide base pair in at least 1% of the population. In contrast, mutations are alterations in DNA that affect less than 1% of the population. SNPs may or may not be associated with any noticeable effects of the individual, whereas the word *mutation* is often associated with a negative effect such as increased risk of disease. In the vast majority of cases, the presence of a SNP or a mutation does not guarantee that a person will get a certain disease and the absence of the SNP or mutation does not guarantee protection against the disease.

Genome-wide association studies compare SNPs between two groups to determine whether there are any substantial differences. SNPs that occur more frequently in one group than the other may help scientists pinpoint genes involved disease, more specifically characterize the type of disease (eg, aggressive or slowly progressive), or identify a group of people who are more responsive to a certain treatment. Still other SNPs can help determine whether someone should be given a lower dose of drug to avoid toxicity.

As more SNPs are discovered and characterized, SNP testing will increasingly make its way into medical care. These tests will likely be designed to help physicians better characterize and treat serious diseases such as cancer. In this way, patients can receive treatments that have the highest likelihood of success for their particular cancer, while minimizing the potential for serious negative effects. This is the type of treatment we all hope for in the future and are even beginning to see glimpses of in medical care today.



## References

Agency for Healthcare Quality and Research (AHRQ). Multigene panels in prostate cancer risk assessment. Available at: <http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1172&pageaction=displayproduct>. Accessed January 25, 2013.

Erichsen HC, Chanock SJ. SNPs in cancer research and treatment. *British Journal of Cancer*. 2004;90(4):747-751.  
Helwick C. 2011 ASCO Annual Meeting: Genes may predict taxane-related peripheral neuropathy. Available at: <http://www.ascopost.com/articles/june-15-2011/genes-may-predict-taxane-related-peripheral-neuropathy/>. Accessed November 19, 2011.

Lindstrom S, Schumacher F, Siddiq A, Travis RC, Campa D, Berndt SI, Diver WR, Severi G, Allen N, Andriole G, Bueno-de-Mesquita B, Chanock SJ, Crawford D, Gaziano JM, Giles GG, Giovannucci E, Guo C, Haiman CA, Hayes RB, Halkjaer J, Hunter DJ, Johansson M, Kaaks R, Kolonel LN, Navarro C, Riboli E, Sacerdote C, Stampfer M, Stram DO, Thun MJ, Trichopoulos D, Virtamo J, Weinstein SJ, Yeager M, Henderson B, Ma J, Le Marchand L, Albanes D, Kraft P. Characterizing associations and SNP-environment interactions for GWAS-identified prostate cancer risk markers—results from BPC3. *PLoS One*. 2011 Feb 24;6(2):e17142.

National Cancer Institute. Understanding cancer series. Available at: <http://www.cancer.gov/cancertopics/understandingcancer/geneticvariation/AllPages>. Accessed October 31, 2011.

National Human Genome Research Institute. Genome wide association studies. Available at: <http://www.genome.gov/20019523>. Accessed November 1, 2011.

Schneider BP, Li L, Miller K, Flockhart D, Radovich M, Hancock BA, Kassem N, Foroud T, Koller DL, Badve SS, Li Z, Partridge AH, O'Neill AM, Sparano AJ, Dang CT, Northfelt DW, Smith ML, Railey E, Sledge GW. Genetic associations with taxane-induced neuropathy by a genome-wide association study (GWAS) in E5103. *J Clin Oncol*. 2011;29: (suppl; abstr 1000).

Schneider BP, Wang M, Radovich M, Sledge GW, Badve S, Thor A, Flockhart DA, Hancock B, Davidson N, Gralow J, Dickler M, Perez EA, Cobleigh M, Shenkier T, Edgerton S, Miller KD; ECOG 2100. Association of vascular endothelial growth factor and vascular endothelial growth factor receptor-2 genetic polymorphisms with outcome in a trial of paclitaxel compared with paclitaxel plus bevacizumab in advanced breast cancer: ECOG 2100. *J Clin Oncol*. 2008;26(28):4672-8.

United States Department of Energy Office of Science. Human Genome Project Information. Available at: [http://www.ornl.gov/sci/techresources/Human\\_Genome/faq/genenumber.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/faq/genenumber.shtml). Accessed October 31, 2011.

United States National Institutes of Health. What are genome wide association studies? Available at: <http://ghr.nlm.nih.gov/handbook/genomicresearch/gwastudies>. Accessed October 31, 2011.

## Why was this guide developed?

As advocates try to work within the system to advance research it is important to understand the basic tenets of the science. By gaining a better understanding, advocates can identify and illustrate the issues and problem-solve to support solutions. The emerging science and issues in research involving biomarkers and genome-wide association studies were the motivation for developing this document. We hope that this information will be helpful to advocates and others interested in advancing the science and improving care for cancer patients.

## About Research Advocacy Network

Research Advocacy Network is committed to improving patient care through research. Our goals are to get results of research studies for new treatments and improved methods of detection of cancer to patients more quickly, to give those touched by the disease an opportunity to give back and to help the medical community improve the design of its research to be more attractive to potential participants. Because research holds the hope for improvements in treatment, diagnostics and prevention, we are dedicated to patient focused research. We believe dissemination of research results to the medical community and patients can have a major impact on clinical practice.

The Research Advocacy Network (RAN) is a not-for-profit, 501(c)(3) tax-exempt organization that was formed in 2003 to bring together participants in the research process with the focus on educating, supporting, and connecting patient advocates with the medical research community. While there are many organizations addressing the needs of patients with specific diseases, political advocacy, cancer education and fundraising, no organization has focused on advancing research through advocacy. RAN works with advocates and organizations to effectively integrate advocates into research activities. Please learn more about us at our Web site at [www.researchadvocacy.org](http://www.researchadvocacy.org) or contact us about our work by e-mailing us at [info@researchadvocacy.org](mailto:info@researchadvocacy.org) or by phone 877-276-2187 or FAX at 888-466-8803.

## Funding

Funding for the development and printing of this material is part of the Patient Advocacy and Care Translation (PACT) Core of the Komen Promise Grant “Comprehensive Biomarker Discovery Project for Bevacizumab in Breast Cancer” at Indiana University Melvin and Bren Simon Cancer Center, Bryan Schneider, M.D., Principal Investigator.

### Reviewers

- Cheryl Jernigan  
Patient Advocate, Susan G. Komen for the Cure® Advocates in Science
- Bryan P. Schneider M.D.  
Associate Professor  
Departments of Medicine & Medical and Molecular Genetics Divisions of  
Hematology/Oncology & Clinical Pharmacology Indiana University School of Medicine

### Development Staff and Contributors

- Nancy Biddle  
Graphic Designer
- Mary Ann Chapman, Ph.D.  
Medical Writer
- Mary Lou Smith  
Co-Founder  
Research Advocacy Network, PACT Core Advocate
- Elda Railey  
Co-Founder  
Research Advocacy Network, PACT Core Advocate