

Applying Preventive Genomic Medicine in Clinical Practice

Navigenics is committed to providing genetic risk information that is fully vetted, transparent and understandable. The results of our service allow physicians to implement probabilistic, pre-symptomatic genetic risk assessment so that behavior modification, screening and early detection can be offered to patients in a personalized manner.

The goal of this document is to provide sufficient background on our service to enable you to make preventive genomic medicine and pharmacogenomics a reality in your practice. The document is organized in two sections and addresses the following key points:

Section 1: Adding genetic analysis to your clinical practice

- Genetic analysis is another tool in clinical decision making
- Individual genetic risk factors have effects equal to and often surpassing those of environmental risk factors and family history
- Combining genetic risk factors can identify individuals with more than ten-fold relative risk of disease compared to the genetic risk in the general population
- Identifying specific genetic variants can inform medication selection
- Certain types of patients may derive particular benefit from the service
- Genetic Counselors give personalized attention to you and your patient

Section 2: The scientific foundation of our service

- Overview of human genome and variation
- Traditional genetic testing is indication-specific and applicable to a small number of individuals
- New generation of genetic testing leverages common genetic variants and is applicable to a large number of individuals
- Strict curation criteria of the genetic risk factors ensures scientific rigor
- Risk model based on genetic information provides discriminatory power to assess predisposition to disease

Section I: Adding genetic testing to your clinical practice

The movement towards personalized medicine and prevention in healthcare

According to the Office of the Surgeon General, seven of ten Americans die each year from a preventable chronic disease such as heart disease, diabetes and many forms of cancer. In addition to the high morbidity of chronic disease, there is an associated high economic cost. For instance, health care spending in the United States reached \$2.3 trillion last year, which is 16% of the country's gross domestic product. As much of this cost is due to treating chronic disease, the Office of the Surgeon General has declared that disease prevention is its number one priority. In addition to focusing their efforts on prevention, government agencies have begun to concentrate on personalized medicine. For example, the Department of Health and Human Services recently created the Personalized Health Care Initiative which aims to improve the safety, quality, and effectiveness of healthcare for every patient. One of the goals of this program is to leverage genetics to support personalized health care so that individuals can be proactive, rather than reactive, in regards to their wellness.

The Navigenics services provide insight into patients' health risks by determining their individual genetic load for a variety of conditions and medication sensitivities where genetic risk factors are well established. The conditions included in Navigenics' analysis are those that are clinically actionable and those that contribute to the major burden of disease in the United States, such as myocardial infarction, cancer, and type 2 diabetes. Each of these health conditions, as well as additional conditions included in the analysis, are caused by both genetic and non-genetic (behavioral, lifestyle, environment) risk factors. By providing information about the level of individual genetic risk, your patients can proactively modify their behavior and reduce their overall disease burden. You can also utilize pharmacogenomic information to optimize drug effectiveness and reduce the risk of side effects.

Genetic testing provides an additional tool in clinical decision making

Genetic factors and environmental factors - lifestyle, personal and family medical history - need to be considered for a complete picture of risk assessment. This is because each of these types of risk factors has advantages and limitations in informing risk. For example, family medical history can capture general risk information for genetic factors inherited from one's parents, but can be incomplete, unknown, or recalled inaccurately. Indeed, family medical history can often be one-dimensional and is affected by family size. Furthermore, family medical history is a summation of both genetic and environment factors, making it sometimes difficult to tease those factors apart.

To illustrate the incompleteness of family history information in capturing common genetic risk for common disease, consider the percent of cancer cases without a family history in the Cancer Genetic Markers of Susceptibility (CGEMS) project. The CGEMS project was launched in early 2006 and is a large, comprehensive study funded by the National Cancer Institute with a goal to identify genetic risk factors for breast and prostate cancer (cgems.cancer.gov). Of 1143 breast cancer cases in the study, only 273 (23%) had a positive family history of the disease. For prostate cancer, only 11% (125 of 1104 prostate cancer cases) had a positive family history.

Environmental and lifestyle risk factors provide a piece of the risk assessment puzzle as well, but since they change over time and are often difficult to measure, they don't provide all of the information. Lastly, genetic risk factors can be measured accurately and do not change over time. Since genetic data is personalized, this kind of risk assessment may provide additional incentive to change behavior.

As more genetic risk factors are being identified, it is clear that their effects on disease risk are just as significant as the effects of environment and family history. To illustrate, cardiovascular diseases are the leading causes of morbidity and mortality in the United States. In particular, years of medical research have identified a number of well-known environmental risk factors for myocardial infarction, including lipids, blood pressure, obesity, smoking, diabetes, and family history. Recently, genetic researchers have identified two genetic risk markers: 9p21 and MTHFD1L (which are both tested in Navigenics' Health Compass service). Each of these risk factors is an important risk factor for myocardial infarction, with effect estimates ranging from 1.5 to 2.0. Homozygotes for the 9p21 marker have an odds ratio of 1.72 (relative risk 1.68) while those with the MTHFD1L marker have an odds ratio of 1.53 (relative risk 1.50).

Understanding the difference between odds ratio and relative risk

To quantify the effect that a variant has on disease two measures are reported in the scientific literature: odds ratio (OR) and relative risk (RR). The two are generally interchangeable; unless the effect size is very large or the disease is very common, both of these situations are usually uncommon in complex human diseases. Each metric can be calculated from a 2x2 table which shows the counts of cases and controls with risk and non risk alleles.

	Case	Control	
Risk allele present	A	B	A+B
Risk allele not present	C	D	C+D
	A+C	B+D	N

The odds ratio is the odds of disease when the risk allele is present divided by the odds of disease when the risk allele is absent. It is used retrospectively as in case/control association studies.

$$OR = \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{A}{B} * \frac{D}{C} = \frac{AD}{BC}$$

The relative risk is the probability of disease when the risk allele is present divided by the probability of disease when the risk allele is absent. It is used prospectively as in studies of population cohorts.

$$\frac{\frac{A}{A+B}}{\frac{C}{C+D}} = \frac{A}{A+B} * \frac{C+D}{C} = \frac{A(C+D)}{C(A+B)}$$

To put the odds ratios of these genetic markers in context, let's consider the effect size of the above mentioned environmental risk factors that physicians currently use to assess patients' likelihood of myocardial infarction. The effect size of the genetic markers 9p21 and MTHFD1L equals or surpasses the effect size of most of the currently recognized medical risk factors -- an insight which many physicians may find illuminating. Furthermore, the Navigenics risk score combines the effect of multiple independent genetic risk factors which leads to even more discrimination of risk from non-risk. In the case of myocardial infarction, if an individual is in the highest percentile (homozygous for both risk factors) the relative risk is more than 1.7 fold compared to the average genetic risk in the population.

Condition	Type of Risk Factor	Risk Factor	Category	Relative Risk*
Myocardial Infarction	Environmental	Blood pressure	Systolic ≥ 160 mmHg or diastolic ≥ 100 mmHg	1.92
	Environmental	LDL-cholesterol	≥ 160 mg/dl	1.74
	Environmental	Smoking	Regular smoker previous 12 months	1.71
	Genetic	9p21	2 risk alleles	1.68 (1.72)
	Genetic/Environmental	Family history	Parent with MI <50 years	1.52
	Genetic	MTHFD1L	2 risk alleles	1.50 (1.53)
	Environmental	Diabetes	Yes	1.47
	Environmental	HDL cholesterol	<35 mg/dl	1.46

*To compare the effect of different types of risk factors, odds ratios (typically reported for genetic risk factors) have been converted into relative risks (typically reported for environmental risk factors) using standard calculations. When this conversion has been performed, the original odds ratios are indicated in parentheses “()”.

Like myocardial infarction, many common diseases have this pattern of similar effect sizes from environmental and genetic factors. An additional two conditions are shown for illustration.

Condition	Type of Risk Factor	Risk Factor	Category	Relative Risk*
Colorectal Cancer	Genetic/Environmental	Family history	First degree relative with colorectal cancer	2.24
	Environmental	Body mass index	BMI > 30 kg/m ²	1.75
	Genetic	CRAC1	2 risk alleles	1.70 (1.70)
	Genetic	8q24	2 risk alleles	1.47 (1.47)
	Genetic	EIF3H	2 risk alleles	1.43 (1.43)
	Genetic	SMAD7	2 risk alleles	1.37 (1.37)
	Environmental	Smoking	Current	1.32
Prostate Cancer	Genetic	11q23	2 risk alleles	1.18 (1.18)
	Genetic	8q24_R2	2 risk alleles	3.12 (3.20)
	Genetic/Environmental	Family history	Brother with prostate cancer	2.87
	Genetic	8q24_R1	2 risk alleles	2.20 (2.23)
	Genetic	8q24_R3	2 risk alleles	1.57 (1.58)
Genetic	17q24.3	2 risk alleles	1.44 (1.45)	

*To compare the effect of different types of risk factors, odds ratios (typically reported for genetic risk factors) have been converted into relative risks (typically reported for environmental risk factors) using standard calculations. When this conversion has been performed, the original odds ratios are indicated in parentheses “()”.

Pharmacogenomic testing allows individualized drug selection and dosing

As efforts to support preventive medicine evolve, so does the need to develop novel approaches to move away from the one-drug-fits-all model. The growing field of pharmacogenomic (PGx) testing allows physicians to optimize drug selection and dosing based on a patient's unique genetic makeup by maximizing effectiveness and minimizing adverse reactions. The application of PGx in clinical practice is expected to improve health outcomes by decreasing medical costs and increasing patient compliance with medication regimens.

Certain aspects of PGx have been around for decades. For example, we learned through trial and error of drug dosing and adverse reactions that patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency should avoid certain medications due to the risk of hemolytic anemia. In recent years a great deal of the groundwork has been performed around chemotherapeutic agents, but the field is expanding to medications that apply to primary clinical practice, including statins, warfarin, and clopidogrel.

Today most physicians believe personal genomic information can be useful in patient care and decision-making, and 98% believe genetic profiles will influence drug therapy. PGx can help you optimize drug selection and dosing based on a patient's unique genetic makeup by:

- Focusing treatment by pre-identifying responders
- Reducing adverse events by predicting populations at risk
- Tailoring dosage to individual genotype
- Identifying orphan drugs for subpopulations

PGx testing can also reveal individual differences in pharmacokinetics, including absorption or distribution,

metabolism and excretion, and can be especially helpful in cases where the prescribed drug has a narrow therapeutic window, such as warfarin.

Clinical studies indicate a wide range of benefits from PGx testing:

- The first large prospective study evaluating the clinical utility of pharmacogenetic markers shows a reduction in hospitalization rates by nearly a third when genetics is used to tailor warfarin therapy.
- Patients carrying at least one CYP2C19*2 allele taking Plavix have elevated platelet reactivity which was associated with poor clinical outcome after coronary stent placement.
- Beta blocker use can improve health outcomes after heart failure among PGx-identified responders.
- Patients with specific genetic variations are less likely to see decreases in cholesterol following statin therapy.
- The FDA has added black box warnings to numerous medications stressing the importance of genetic information in determining patient benefit.
- Genetic polymorphisms explain 40 percent of the variability in warfarin dose needed for a stable INR, which is more than two times the contribution of age and weight combined.

For the drug interactions covered in our service, personalized pharmacogenomic results reveal your patient's risks of serious side effects or reduced drug effectiveness, allowing you to make clinical decisions that reduce risks and improve therapeutic outcomes before treatment starts.

Finding the ideal patients to target for genetic testing

The Navigenics services are most effective when targeted to those patients who stand to benefit most from the results. There are five main groups of patients that physicians have found benefit in testing:

1. Patients who require additional motivation to adhere to primary prevention strategies such as for cardiovascular or metabolic diseases
2. Patients seeking enhanced personalized screening programs such as for cancer or ophthalmic disease
3. Patients taking or considering initiating medications where pharmacogenomic insights can help maximize effectiveness and minimize side effects
4. Patients with difficult-to-diagnose diseases like Celiac disease
5. Patients with limited or missing family history information

Patients who require additional motivation for primary prevention

Motivation and compliance are difficult areas in the practice of medicine. Genetic testing provides patients with risk scores for conditions that can improve their motivation and compliance with a medical regimen or health behavior changes. For instance, patients who see they are at increased risk of myocardial infarction or type 2 diabetes may better adhere to exercise, diet, and medication regimens. Indeed, some early studies have revealed that giving patients' insight into their genome can have positive behavioral effects. The REVEAL Study (J Geriatric Psychiatry Neurol, 18:250. 2005, Alzheimer Dis Assoc Disorder 22:94 2008) found that participants who learned their genetic risk for Alzheimer's significantly changed their lifestyle behaviors for the better. Another recently published study examined behavior changes in 59 individuals with a family history of melanoma after genetic testing for CDKN2A/p16 mutations, which is known to be mutated in familial melanoma (Cancer Epi Biomarkers and Prevention 17:1510. 2008). The study found improved screening behaviors in mutation carriers and no evidence of a false sense of security among non-carriers. Importantly, the study found that the results of genetic testing improved compliance with suggested screening beyond that obtained with counseling based on family history alone.

Additionally, patients have also expressed appreciation for the fact that understanding their genetic risks allows them to focus on behaviors that are most relevant for them. The modern day era of public health messaging has a multitude of recommendations for all aspects of health that patients may find overwhelming and consequently tune out. Demonstrating to patients that they may be at increased risk for one or two

conditions allows them a more finite and concrete set of health actions to pursue.

Patients seeking enhanced personalized screening programs

Current preventive screening guidelines are based upon criteria that maximize the health of a population but may not always take individual patient differences into account. For instance, mammograms are recommended to begin at age 40 and colonoscopies at age 50, regardless of environmental or individual common genetic risk. Consequently, there are individuals who, based upon their genetic predisposition, could benefit from earlier or more personalized screening regimens. For example, a 39 year old female with markedly increased risk for colon cancer may not want to wait until age 50 to begin screening colonoscopies. A 40 year old male with increased risk for macular degeneration or glaucoma should consider annual ophthalmic exams. In both examples, earlier diagnosis may lead to better disease outcomes.

Patients taking or considering specific medications

The growing field of pharmacogenomic testing enables individualized drug selection and dosing based on a patient's unique genotypes to move beyond the one-size-fits-all model. Pharmacogenomic data allows physicians to move beyond traditional risk factors and considerations, such as comorbidities and comedications, in order to optimize drug response. Where genetic variations impact a drug's metabolic pathway – such as in the active or prodrug state – provides novel insight to pre-identify non-responders, those who are likely to experience adverse reactions, and for some medications, tailor dosing. While the majority of early research focused on chemotherapeutic agents, the field continues to expand to medications that apply to other specialties and primary clinical practice, including statins, warfarin, and clopidogrel.

Patients who have diseases that may be difficult to diagnose

Genetic testing can identify predisposition in patients that can help focus the physician's efforts, particularly in difficult-to-diagnose cases. Some diseases such as celiac disease are challenging to diagnose given their vague and protean symptoms. Navigenics services can identify the likelihood of these diseases in an individual patient as compared to the overall population. Physicians have found this type of information useful when considering differential diagnoses. For example, there have now been cases of patients with intermittent abdominal pain and bloating for more than 5 years where the Navigenics service prompted the physician to order a tTG and intestinal biopsy, which led to a celiac diagnosis.

Patients with limited family histories

Many patients have limited insight into their family history due to adoption, conception techniques such as sperm donation or *in vitro* fertilization using an egg donor, poor medical records, or cultural barriers. An estimated five million Americans are adoptees, for example, with an estimated 125,000 babies newly adopted in the United States each year. Tens of thousands of babies are conceived with the help of sperm or egg donors each year, and many of those donors are anonymous. Many other individuals know the identity of their biological parents, but may not have contact with their biological mother or father. Consequently, one of the useful tools of risk assessment is unavailable to physicians when counseling their patients. The Navigenics service can be used to gain insight into their genetic inheritance for medical purposes.

Genetic counselors facilitate your practice of preventive genomic medicine

The Navigenics service includes unlimited access to our team of Genetic Counselors for both physicians and patients. Genetic counselors are master-level trained specialists, certified by the American Board of Genetic Counseling, with expertise in medical genetics, risk communication, and strategies for helping people manage genetic information. In practice, genetic counselors work with health care providers and patients to identify hereditary risks, discuss options in genetic testing, explain test results and review available diagnostic, preventive and early detection options. Navigenics Genetic Counselors work with both patients and physicians, helping them to understand genome wide test results, and facilitating informed clinical decision-making and medical management. For instance, physicians frequently call the Navigenics counselors to discuss their patients' test results, ask questions about genomics, and explore commonly asked patient questions. The Genetic Counselors are also available for onsite visits to train physicians' office staff.

Section II:

The scientific foundation of our service

Overview of human genome and variation

The human genome comprises three billion bases organized into 23 chromosomes and approximately 25,000 genes. Each human individual carries two copies of each chromosome. There are four types of bases (or nucleotides): A (adenine), C (cytosine), G (guanine), and T (thymine) arranged in various ways to determine any DNA sequence. Some sequences, known as “coding regions”, are translated into RNA and subsequently into proteins. Other sections of DNA, approximately 95% of the genome, do not become proteins. At one time, these non-coding sequences were known as “junk DNA”, but there is a growing understanding that non-coding regions play an important role that is just not yet fully understood.

If two DNA sequences are compared, either in one individual or between two individuals, the bases would be about 99 percent identical. Variations in DNA sequence can involve large stretches of nucleotides (chromosomal rearrangements, or insertion or deletion of repetitive elements), multiple nucleotides (called microsatellites, or variable number tandem repeats), or even a single base (single nucleotide polymorphisms, or SNPs, pronounced “snips”). This publication will focus on SNPs since they are the most common form of variation and are typically used in the new generation of genetic testing.

SNPs are among the simplest type of genetic variation to study in the laboratory because they are binary, involve a single location, and do not involve repetitive elements. A SNP is defined as a single base in the human genome that differs among individuals in the population. Historically, the difference is quantified in the following way: if one of these genetic markers is present in 1 percent or more of the population then the variation is a “polymorphism.” If it is present in less than 1 percent it is a “mutation.” However, this strict distinction between disease-causing mutations and normal polymorphisms based on frequency is arbitrary; rather, it is more accurate to conceptualize that locations in the genome vary among individuals and that some of these variations are disease-causing. A location in the genome that is a SNP will usually have just two variants, also called alleles, present in the population – for instance, at a given SNP, each chromosome will have either an A or a T. Since an individual has two copies of each chromosome, there are three possible genotype combinations: AA, AT, or TT. SNPs are located throughout the genome, in both the coding and non-coding regions. SNPs in non-coding regions can produce phenotypes (the outward expression of a DNA change, or genotype) by altering the expression of a gene or they may result in no discernible phenotype. SNPs in coding regions can alter the function of a protein, by preventing it from being produced correctly, or have no functional affect. Approximately half of the 10 million SNPs that are thought to exist in the human genome have been cataloged by researchers into the dbSNP public database (www.ncbi.nlm.nih.gov/projects/SNP). The International HapMap project (www.hapmap.org) has further explored the frequency of 3 million SNP alleles in multiple diverse populations.

Genotyping technologies

Since the great majority of the human genome does not vary between or within an individual, it is only necessary to examine targeted parts of the genome to determine the bases that an individual has at a particular SNP. This process of determining alleles at particular non-contiguous DNA positions is called genotyping. When the alleles are determined at contiguous DNA positions the laboratory process is called sequencing. Sequencing can refer to a specific gene, region, or entire genome. SNPs are genotyped in the laboratory using any one of numerous technologies. The choice of genotyping technology is determined by several logistic and operational considerations, rather than technical performance. Navigenics uses multiple platforms in order to capture as much relevant genetic information as possible.

Traditional genetic testing focuses on rare variants

Traditionally, genetic testing in the clinic has focused on uncommon, familial conditions caused by rare DNA mutations that are highly penetrant (that is, most individuals that carry the mutation will also be afflicted with the condition) or on conditions that are common in particular ethnic groups. For example, familial hypercholesterolemia is caused by more than 1,300 DNA mutations in the LDLR or APOB genes. Such conditions are often called Mendelian conditions, because of their predictable patterns of inheritance. Genetic testing for this type of condition is usually performed to confirm a diagnosis and to identify other family members that might be at risk or for family planning purposes. The second type of traditional genetic

testing is called carrier screening. The primary purpose of carrier screening is for family planning since while having one copy, i.e. a carrier, of the disease mutation is not in and of itself harmful, carriers who breed with another carrier are at risk of having children with two copies, which leads to symptomatic disease. Carrier screening is often directed at the at-risk population, such as Caucasians for cystic fibrosis or Ashkenazi Jews for Tay-Sachs disease. Sometimes, a particular genetic test may not fall neatly into both categories. For instance, BRCA1 and BRCA2 gene mutations are usually only found in rare, familial forms of breast cancer (less than 5-10 percent of all breast cancer cases) or in certain Ashkenazi Jewish populations, so these are the populations where the test is indicated. In summary, traditional genetic testing for conditions is indication-specific, and is typically informed by personal, family, and ethnic ancestry information.

Identifying genetic risk factors for complex human disease

Complex diseases, which do not follow a predictable Mendelian pattern of inheritance, have been shown to be affected by both genetics and environment. For example, the genetic contribution to type 2 diabetes can be estimated by studying both identical and fraternal twins. If each individual in the twin pair is diagnosed with type 2 diabetes then they are concordant; if only one suffers from the disease then the twins are called discordant. The comparison of concordance rate between identical twins, who share 100 percent of their DNA, with the concordance rate between fraternal twins, who share 50 percent of their DNA, enables researchers to estimate the genetic and non-genetic contribution to the disease. The genetic contribution measured in these studies is termed heritability. The heritability of complex human disease is typically around 50 percent, indicating that a significant fraction of disease risk is both genetic and environmental. This is important because it indicates that some of the disease risk can be modified by changing an individuals' environment.

In the past ten years there has been remarkable progress in identifying the genetic variants that contribute to complex disease. This progress has been made possible by the Human Genome Project, which resulted in the complete human genome sequence. This project advanced the field of genetics and genomics by providing a map of all the genes and their relationship to each other, millions of genetic markers that could be used for research, and scalable laboratory operations. The International HapMap project has further facilitated genetic research by characterizing the pattern of human variation and validating millions of genetic markers that commercial companies are using to make high-throughput genotyping products. These advances enabled genetics researchers to use association analysis (where an increase in allele frequency in a population of cases as opposed to a population of controls is observed) rather than linkage (where a segment of DNA is inherited in families more often than expected) to map disease variants. Association, compared to linkage analysis, is a more powerful method to identify low penetrant variants that are common in the general population and predispose individuals to common disease.

Association tests have been used for many years, starting with the first associations of disease with the ABO blood locus in 1953 (Bodmer and Bonilla. *Nature Genetics* 40:695, 2008), followed closely by numerous associations to the HLA locus on chromosome 6. Historically, association testing was limited to genetic variants in candidate genes that made "biological" sense. For instance, candidate genes for lupus might be genes involved in regulating the immune response. With the completion of the Human Genome and HapMap projects and the subsequent high-throughput genotyping products that ensued, researchers could query the entire genome without the need for selecting candidate genes or variants. Since 2005, over 150 genome-wide association studies have been performed and published (www.genome.gov/26525384), greatly increasing our knowledge of common genetic risk factors that contribute to complex disease.

What has been learned about the genetic architecture of complex disease?

The great majority of variants that have been associated with common disease are usually of relatively low effect size (1.5 to 2.0 odds ratio per allele). Secondly, there are multiple variants that contribute to disease, and they appear to act independently of each other. By combining the genetic risk across multiple low-effect-size loci, the overall risk of disease can be greatly increased. (For example, see Lu and Elston, *AJHG* 82:641. 2008 for type 2 diabetes, Morisson *et al.* *American Journal of Epidemiology* 166:28. 2007 for coronary heart disease, and Gold *et al.* *Nature Genetics* 38:458. 2006 for age-related macular degeneration.) Thirdly, the

risk allele is often the higher frequency allele (this is in direct contrast with Mendelian mutations, where the risk allele is almost always the rare allele). Finally, the association is often seen in other populations with different ancestries, although relatively few investigators have specifically studied ethnic groups outside of the Caucasian ethnicity. When considering the transferability of genetic associations studied in one ancestral group compared to another it is important to consider both the variation in effect size and the variation in risk allele frequency. Loannidis *et al.* (Nature Genetics 36:1312, 2004) examined 43 genetic associations studied in 697 populations with different ancestries. They show that the majority (58%) of loci show significant variation in the frequency of the risk marker, but that a minority (14%) of loci show statistically significant, but often minor, heterogeneity in the effect size of the risk marker. This means that the majority of genetic associations can be applied to other non-studied populations as long as the frequency of the variant is taken into account. This transferability of common risk alleles for common disease is again in contrast to Mendelian mutations, which are sometimes seen only in isolated families or ethnic groups (presumably due to founder population effects or inbreeding).

The previous generation of genetic testing measured gene variations that were highly penetrant, but were very rare. Consequently, this kind of genetic medicine was relatively diagnostic, but was applicable to a small minority of individuals. In contrast, the new generation of genetic testing measures gene variants that are less penetrant, but relatively common. Thus, these genetic variants are of interest to many more individuals, but are more probabilistic in their predictive value.

Traditional Genetic Testing	New Generation Genetic Testing
Rare, highly penetrant mutations	Common, low penetrant variations
Highly predictable disease outcome	Probabilistic disease outcome
Specific, predictable impact on family inheritance	Doesn't yield predictable family inheritance patterns
Public health impact is low	Public health impact is high
Individual impact is high	Individual impact is low to high depending on genotypes
	Interaction with environmental risk factors

Risk score model based on strict curation criteria creates a high scientific rigor

Navigenics has developed a method to combine genetic risk information across different risk markers to produce a composite genetic risk score. The foundation for the risk score is the individual genetic risk factors that are associated with a condition. We have developed a set of criteria used to evaluate genetic associations that are reported in the scientific literature:

- We require that an association be consistently replicated within a given ethnic group. For candidate gene studies we require at least two independent replications (since publication of such studies can be difficult when results are negative, publication bias is a major concern). We further evaluate the entire body of literature and require that at least 60 percent of published reports show consistent association.
- We take a close look at sample size and have a minimum requirement of 250 cases and 250 controls when the effect size is less than 1.5. However, most of the associations that we report on have been studied in several thousand individuals.
- We evaluate the study design to determine how the cases and controls were selected, phenotyped, and genotyped.
- We critically evaluate the significance level of each reported SNP, taking into account the stricter significance levels needed when multiple testing is performed.

Navigenics employs these strict criteria before including a SNP in our risk score model. This is because

for many years genetic association studies have had a poor reputation. Specifically, the literature is littered with genetic associations that were observed once but could then never be replicated in subsequent studies. A review of 166 gene-disease associations (which included 633 studies) published between 1986 and 2000 found that only 6 were consistently observed in multiple, independent studies (Hirschorn *et al.* Genet Med. 4:45, 2002). The poor performance of early association studies has been attributed to multiple factors (Cardon and Bell. Nature Review Genetics 2:91, 2001), which include small sample size, subgroup analysis and multiple hypothesis testing, poorly selected control group, failure to attempt study replication, failure to detect linkage disequilibrium with adjacent loci, over-interpreting results, positive publication bias, inadequate attention to genotyping assay and algorithm quality, and failure to correct or avoid population substructure. The failure to correct for population substructure (i.e. genetic differences in ethnic groups) provides an especially illustrative example in the challenge of conducting these studies. Historically, disease cases and controls were not well matched for ethnicity. Consequently, genetic differences between the case and control groups were actually a function of the different ethnicities and not a function of disease-causing genes.

Recent genome-wide association studies (GWAS) have made noticeable improvements in addressing these failure points; however, it is still necessary to review these articles with a critical eye. For instance, although the great majority of recent studies have analyzed large samples sizes and often include multiple replications in multiple population samples, one still needs to evaluate the overall study design, control selection and statistical correction for testing multiple hypotheses. Furthermore, with multiple independent studies being published, often using different genetic markers, it is necessary for subject matter experts to carefully examine the overall evidence to come to a comprehensive conclusion.

Even with these guidelines for curation, interpretation of the genetic association literature is challenging. This is why Navigenics has hired a team of subject matter experts that carefully reviews each genetic risk factor that is included in our service. Only about 5-10% of papers reported in the scientific literature pass our strict criteria. You can find more information about the conditions that are not included in our service and why on our website (www.navigenics.com).

Analytic accuracy of the genotyping result is >99%

After careful curation of the literature, we determine if the associated SNP can be tested, either directly or indirectly, on the Affy6.0 genotyping platform. A direct test is possible when the exact published SNP is represented on the array. An indirect test is used when the exact published SNP is not on the array, but a "tag" SNP is present. Tag SNPs are proxies for the published SNP that are identified by analyzing the linkage disequilibrium pattern in that region of the genome. Linkage disequilibrium is a genetic concept that describes the co-inheritance of chunks of DNA. When a whole segment of DNA is in linkage disequilibrium, SNP variants will tend to be inherited together and can therefore serve as proxies for each other. The results of the HapMap project (www.hapmap.org) provide the raw material for this sophisticated analysis of linkage disequilibrium patterns and the determination of appropriate proxy, or tag SNPs. When a published SNP cannot be accessed directly or indirectly on the Affy6.0 array, Navigenics fills in the gaps using a secondary technology, called TaqMan.

We pay careful attention to the analytic accuracy of our test SNPs. Before including a SNP in the combined risk score, we analyze the raw data from the array in a set of 270 reference samples with known HapMap genotypes. We evaluate the raw data to ensure that the genotype clusters are well separated and are of the expected genotype frequencies. We test the reproducibility and accuracy of the results by running the assay in the reference population multiple times and comparing the genotypes obtained with genotypes from the published database. For individual markers we require an accuracy rate of >99% in the validation study.

Combining the effect of individual genetic risk factors into one composite risk score

Navigenics has developed a model-based risk estimate that combines genetic risk across loci that have been associated with disease. After our rigorous curation procedure, we start with the genotypic odds ratios reported in the literature for each locus affecting a condition. There are two effect size measurements that

are reported in the literature: odds ratios and relative risks. We convert the odds ratios into relative lifetime risks using the prevalence of the condition and the genotype frequency in a reference population (typically the HapMap population comprising individuals from the U.S. with Northern and Western European ancestry). We do this because odds ratios can overestimate disease risk especially when the prevalence of the condition is greater than ten percent or the effect size is larger than two.

Understanding the Navigenics risk score model

The risk model combines information from multiple genetic risk factors and takes into account both the effect size of the markers and their genotype frequencies.

1. For each marker to be included in the risk score, the published odds ratios for each possible genotype (2 risk markers, 1 risk marker, and 0 risk markers) are converted into relative risks:

Marker 1: 10.57, 2.72, 1 → 9.06, 2.64, 1

Marker 2: 9.99, 3.16, 1 → 9.47, 3.12, 1

Marker 3: 6.98, 2.33, 1 → 6.78, 2.31, 1

2. A weighted score for each marker is calculated based on the relative risk and genotype frequencies in the reference population:

Marker 1: $(9.06 * 0.016) + (2.64 * 0.40) + (1.0 * 0.58) = 1.78$

Marker 2: $(9.47 * 0.40) + (3.12 * 0.38) + (1.0 * 0.21) = 5.18$

Marker 3: $(6.78 * 0.88) + (2.31 * 0.12) + (1.0 * 0.00) = 6.24$

3. A multiplicative model is used to calculate a weighted average score in the reference population across all markers:

$1.78 * 5.18 * 6.24 = 57.54$

4. Person X is genotyped for markers associated with the disease and their multiplicative score is calculated. Assume the individual has 1 risk marker at Marker 1 and 2 risk markers for both Markers 2 and 3. Their multiplicative score is:

$2.64 * 9.47 * 6.78 = 169.51$

5. Divide person X's score (169.51) by the average score in the reference population (57.54) to determine the normalized relative risk:

$169.51 / 57.54 = 2.95$

6. Compare person X's score (169.51) to the distribution of scores in the reference population to determine the percentile an individual falls in. In other words, we show how many individuals in the reference population have a score higher, lower, or the same as person X. Individuals with higher scores are more likely to have the disease.

18% of individuals have the same genetic risk

82% of individuals have a lower genetic risk

7. To determine the estimated lifetime risk for person X, we take the fraction of the genetic load ($169.51/57.54$) and multiply by the average gender specific lifetime risk (3.1%).

$(169.51 / 57.54) * 3.1\% = 9\%$

We use a multiplicative model for combining the relative risks across multiple loci. This model produces a score that can be compared to either the average score in the reference population or the distribution of scores in a reference population.

- When a person's score is divided by the average score we have calculated the relative risk of disease compared to the average population risk, or the normalized relative risk.

- When a person's score is compared to the distribution of scores in the reference population we have determined the placement of that individual, or the percentile, in the reference population. Because different conditions have different numbers of genetic risk factors the percentile is expressed as a range with upper and lower boundaries.

Lastly, to determine a person's estimated lifetime risk we take the normalized relative risk and multiply it by the average lifetime risk, which is compiled from authoritative epidemiological reports in the medical literature. The figures are based on the total lifetime risk for that condition, typically for a U.S. population and for that person's gender. The lifetime risk for some of the less common conditions is difficult to know with certainty and is estimated from the best available published evidence, using information about the incidence of the condition and the population age distribution in the United States.

Limitations of the risk model

Our risk score model has some important assumptions that may affect the accuracy of the results. Specifically, the formula assumes that the odds ratios are known, that the prevalence of the condition is known, that the genotype frequencies in the reference population are representative of the individual's own ancestry, and that each marker has an independent effect on the overall risk. There are other important limitations of our test: The result is probabilistic, not diagnostic, and it does not include all genetic risk factors. For example, risk factors that cause monogenic forms of the disease are typically not included. Lastly, researchers are continuing to discover variants that contribute to disease. These genetic updates may change an individual's overall score so Navigenics actively updates our risk models as new literature is vetted by the curation team. As these additional data inputs are incorporated into the Navigenics services, the risk model becomes even more predictive.

Genetic risk is presented using different visualizations to enhance understanding

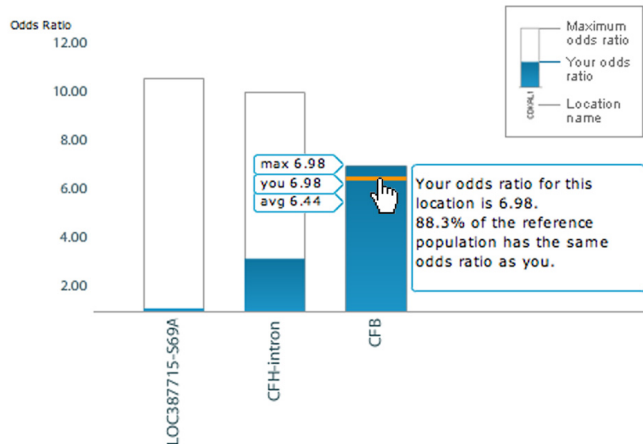
An individual's results are presented through our secure website. In brief, our test result encompasses both the individual's genotype specific odds ratio for each genetic marker and their composite score. The composite score is presented as a percentile range relative to a reference population. We have developed a number of ways to communicate the risk for common diseases to you and your patient.

When a person first signs in, a tutorial walks them through the interpretation of the results from our examination of their DNA. At the overview page, all of the conditions we currently test for are listed and arranged by your patient's estimated lifetime risk percentage. The estimated lifetime risk is the estimated risk of manifesting the condition in an average lifetime based on your patient's gender. Each condition is color-coded gray or orange based on a comparison of the individual's estimated lifetime risk and the average lifetime risk. The average lifetime risk encompasses both average environmental and average genetic risk, while the individual's estimated lifetime risk is the increase or decrease from the average based on the genetic markers we assessed. Orange indicates an overall lifetime risk greater than 25 percent (to call attention to conditions with an elevated absolute risk) or risk that is more than 20 percent above average for that particular condition (to call attention to an increased relative risk for less common conditions). The normalized relative risk (the relative risk of an individual compared to the average genetic risk in the population) can be determined by dividing the individual's estimated lifetime risk by the average lifetime risk in the population.

For each condition the individual can get more detailed information on the genetics and condition information, as well as risk factors, causes, prevention and intervention strategies. In order to give your patient a picture of where they are relative to a reference population, your patient's genetic make-up for the tested SNP is compared to the genotypes of the HapMap reference population. This visualization shows what percent of the reference population has a score lower, the same as, or higher than the individual. Individuals at the lower end of the distribution have low to moderate genetic load and individuals at the very high end of the distribution have a high genetic load (i.e. carry many risk markers). Using computer models we have shown that individuals with moderately increased genetic risk typically have a two-fold increase in odds of disease compared with individuals with less than moderate genetic risk, however individuals at the extreme upper end of the distribution (those in the highest percentile of the distribution) have odds of four to greater than ten depending on the condition considered.

The genotype for each SNP is presented in a table, along with the effect estimate and a link to the publication about the SNP. Additionally, a SNP bar chart shows the relative contribution of each marker tested to the individual's overall estimated lifetime risk. The contribution of a marker is a combination of both the effect estimate and the frequency of that genotype in the population.

Different ways of visualizing and presenting genetic risk information

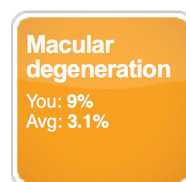
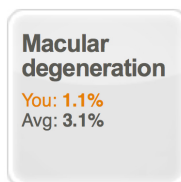
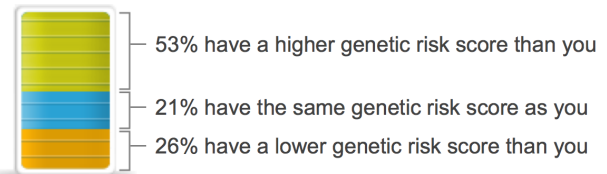


A graphical representation of each marker in the risk assessment provides information about the relative contribution of that marker genotype to the individual's overall score. This chart displays both relative and absolute risk.

Odds ratios specific to the individual's genotype are presented along with a lay summary of the published article.

Gene or location	Risk marker	Your markers	Odds ratio	Source
LOC387715-S69A	T	GG	1.0	American Journal of Human Genetics, 2005
CFH-intron	A	CA	3.16	Nature Genetics, 2007
CFB	T	TT	6.98	Nature Genetics, 2006

A graphical picture of an individual's overall score compared with the reference population. Individuals that have a high percentile, with few reference individuals having a higher score, are at much higher risk of disease than individuals with lower percentiles.



The average lifetime risk of a condition and the individual's estimated lifetime risk based on their genetic risk factors is shown in this condition box. When the estimated lifetime risk is greater than 20% the average or when it is greater than 25% the condition box is orange. The normalized relative risk (the relative risk of an individual compared to the average genetic risk in the population) can be determined by dividing the individual's estimated lifetime risk by the average lifetime risk in the population.

Performance of the Navigenics risk score model

We have evaluated our risk model by calculating receiver operating characteristic (ROC) curves using genotype data from three studies (CGEMS, WTCCC, and AREDS) accessed through dbGAP (www.ncbi.nlm.nih.gov/sites/entrez?db=gap). The CGEMS (Cancer Genetics Markers of Susceptibility) study is a three year initiative of the National Cancer Institute and aims to identify genes that increase or decrease the risk of cancer. The WTCCC, or Wellcome Trust Case Control Consortium, is a collaboration of UK human geneticists interested in identifying common genetic variation that predisposes to a variety of common conditions. Lastly, the AREDS, Age-Related Eye Disease, study is a major clinical trial sponsored by the National Eye Institute that has collected both genotype and phenotype information. We analyzed the CGEMS dataset for prostate cancer; breast cancer; the WTCCC dataset for coronary artery disease, Crohn's disease, rheumatoid arthritis, and type 2 diabetes; and the AREDS dataset for macular degeneration. The AUCs ranged from 0.57 to 0.72. An important caveat of these publicly available datasets is that they do not exactly contain the markers used in the Navigenics risk score. This is because the studies have been

performed using older genotyping technologies which do not contain all the relevant SNPs or because the risk model uses genetic risk markers identified in more than one study. To determine the performance of our test using the exact markers included in the service we used computer models. We modeled a case/control study using 10 million individuals for age related macular degeneration, breast cancer, myocardial infarction, prostate cancer, type 2 diabetes, and 16 other conditions included in our service and observed AUCs ranging from 0.54 to 0.77. Many of the commonly used diagnostics in medicine such as the PSA test, stress treadmill test, and mammography have similar AUCs (ranging from 0.62-0.85). In particular published AUC values for the commonly used stress treadmill test range from 0.62 (Am Heart J 142:136. 2001) to 0.79 (Chest 119:1933. 2001). In summary, AUC analysis using both real and simulation genotype data show that the Navigenics risk score model does provide discrimination between cases and controls even in the absence of family history and environmental risk factor assessment.

A refresher on receiver operating characteristic (ROC) curves

ROC curves have been used as a measure of the reliability of a risk assessment test. For a perfect test, a threshold t would be chosen such that all individuals with a score larger than t would develop the conditions, and all individuals with a score less than t would not. However, in practice we will find that for any given threshold there is some fraction of false positive and false negative assignments. The ROC curve graphically depicts the relationship between false positive rates and true positive rates, and thus it can be used to guide the tradeoffs between test sensitivity and specificity. We use the area under the ROC curve (AUC) as a quantitative measure to compare different risk estimate scores. AUC is a measure of the accuracy of the test, where an AUC=0.5 represents a test that cannot correctly distinguish between those with the disease and those without, and an AUC=1 represents a test that perfectly distinguishes the disease state.

We examined the relationship between high Navigenics risk scores and a positive family history of disease using the aforementioned CGEMS data set. We determined that 76 of the 1143 breast cancer patients fell into the top 10% of genetic load based on our reference population. The majority (70%) of these 76 breast cancer patients with the highest genetic risk did not have a positive family history. Conversely, 289 of the 1104 prostate cancer patients fell into the top 10% of genetic load based on our reference population. The majority (87%) of these 289 prostate cancer patients with the highest genetic risk did not have a positive family history. This means that many individuals with the highest genetic risk would not be discerned by determining family history alone. It is key to note, that family history, genetic testing, and environmental risk assessment are important pieces of a comprehensive risk assessment since the overlap of cases with the highest genetic risk and individuals with a positive family history in the CGEMS data sets is not large.

Conclusion

We hope that this paper has served as an overview of our service, our science, and our test. Navigenics is ready to embark with you on a journey of individualized medicine and scientific discovery that will shift your practice of medicine from personal to personalized. If at any time you would like to learn more, or if you have specific clinical or science-related questions, we encourage you to speak directly with one of our Genetic Counselors by calling (866) 522-1585 in the United States or (650) 585-7744 from international locales.

References

Colorectal Cancer

Genetic risk factors:

Nat Gen 39:954. 2007
Nat Gen 39:1315. 2007
Nat Gen 40:26. 2008
Nat Gen 40:631. 2008
Nat Gen 40:623. 2008

Other risk factors:

Eur J Cancer 42:216. 2006
Am J Epi 152:847. 2000
JNCI 92:23.2000

Myocardial Infarction

Genetic risk factors:

Science 31:1491. 2007
NEJM 357:443. 2007

Other risk factors:

Am J Cardiology 99:1236. 2007
Circulation 97:1837. 1998

Pharmacogenomics

Warfarin

Gen in Med 10: 89. 2008

Prostate Cancer

Genetic risk factors:

Nat Gen 39:645. 2007
Nat Gen 39:631. 2007
Nat Gen 39:977. 2007

Other risk factors:

Int J Cancer 107:797. 2003

AUC measures in commonly used diagnostics

AJR 174:1257. 2000
Urology 62:835. 2003
Am Heart J 142:136. 2001

Notes

An additional page for your notes.