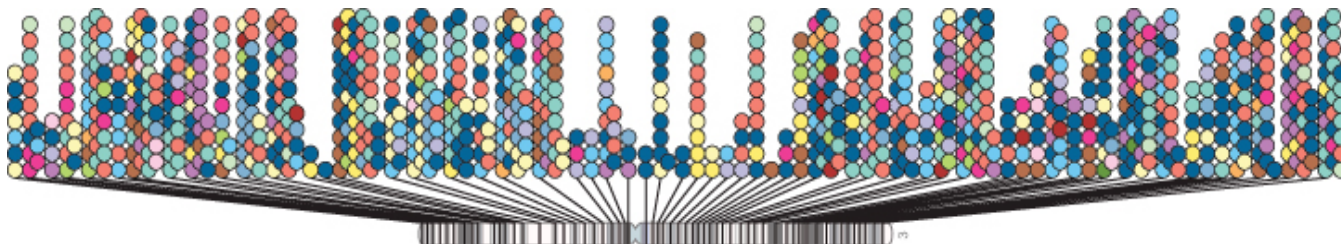


# Living in a Genomic World: Unraveling the Complexity of GWAS Studies

by

Rivka L. Glaser, Department of Biology, Stevenson University, Stevenson, MD

Erin L. Zimmer, Department of Biology, Lewis University, Romeoville, IL



## Part – Introduction

Josie, a sophomore biology major at Tufton University, was currently enrolled in a genetics class. The class had recently been learning about genome-wide association studies (GWAS) and Josie was intrigued by the massive amount of information that could be obtained from such studies. In the last lecture, her professor talked about several companies that offered genomic testing to the general public. Josie was surprised to find out that one of the U.S. companies was shut down by the FDA in November 2013, but then resumed its services in 2015.

Josie's professor assigned the students a poster presentation on how current knowledge of DNA has impacted modern life. Fascinated by GWAS studies and the technology that enables the general public to access their genomes, Josie decided to focus on these topics for the poster presentation. She planned on researching a few of these companies in depth to find out the different traits that each company tests for. Knowing that she had a family history of Alzheimer's disease, Josie went a step further and decided to have her genome sequenced by one of these companies.

### Genome-Wide Association Studies (GWAS)

GWAS studies are rapidly becoming the norm for identifying genes or polymorphisms that may be involved in human diseases and phenotypes. These studies rely on single nucleotide polymorphisms (SNPs) (Figure 1). Although some SNPs are in genes, the majority of SNPs are found in non-coding regions of the genome.

```

Person 1:  A A T A C G T G A G T G A G G C C T T A A
Person 2:  A A T A C G T G A G T G A G G C C T T A A
Person 3:  A A C A C G T G A G T G A G G C C T T A A
Person 4:  A A T A C G T G A G T G A G G C C T T A A
Person 5:  A A T A C G T G A G T G A G G C C T T A A
Person 6:  A A G A C G T G A G T G A G G C C T T A A
Person 7:  A A T A C G T G A G T G A G G C C T T A A
  
```

Figure 1. Example of a SNP. DNA sequences from seven individuals are pictured above. Notice in the third nucleotide of the sequence, Person 3 has a C and Person 6 has a G whereas everyone else has a T at this position.

SNPs are present in more than 1% of the population. All known SNPs are catalogued in a database called dbSNP (<<https://www.ncbi.nlm.nih.gov/snp>>) and given a unique identifier called an *rs* number (e.g., rs1799853). The *rs* number is an accession number used by researchers to refer to a specific SNP.

*Questions*

1. If the human genome is three billion base pairs long, estimate the average number of SNPs in the human genome.
2. Predict the different effects that a SNP would have if it were located in an exon, an intron, or in a intragenic region.

GWAS studies compare the frequency of SNPs in a control population against the frequency of the same SNPs in an affected population. Typical GWAS studies recruit thousands of individuals and look at thousands to millions of genetic variants in each individual. The amount of data generated from such a study is immense and has been termed “big data.” Microarrays, whole genome sequencing, and exome sequencing are all technologies that are used to analyze vast amounts of variants from many individuals.

3. Compare and contrast these three technologies (microarrays, whole genome sequencing, exome sequencing) in terms of their methodologies and the information that they can provide. The following resources may be helpful:
  - Genetic Science Learning Center. 2013. DNA microarray. <<http://learn.genetics.utah.edu/content/labs/microarray/>>
  - My46 (University of Washington). *n.d.* Whole genome and exome sequencing. <<https://www.my46.org/intro/whole-genome-and-exome-sequencing>>
  - NISC. 2015. Whole exome sequencing and analysis. <[http://www.nisc.nih.gov/docs/FAQ\\_whole\\_exome.pdf](http://www.nisc.nih.gov/docs/FAQ_whole_exome.pdf)>

*Microarray:*

*Whole genome sequencing:*

*Exome sequencing:*

4. Direct-to-consumer companies, like 23andMe and Gentle Labs, offer genomic screening to individuals for fees ranging from several hundreds of dollars to several thousands of dollars. The technology used, the number of SNPs tested, and the number of traits for which they screen vary by company. Research these two companies and complete the chart below.

<i>Company</i>	<i>Genotyping technology used</i>	<i>What traits/conditions do they test for?</i>
23andMe		
Gentle Labs		

Gentle for ▾

Why get tested?      Why Gentle?      How it works      Talk to us

## Why Gentle?

You're probably familiar with other DNA testing companies. See how Gentle surpasses the competition by offering more tests, screening more markers per test and by providing free ground-breaking mobile apps.

Company	Percentage of genes sequenced
others	1.9%
gentle	100%

percentage of genes sequenced

### We sequence all your genes. Not just 1.9% of them.

We all know \$99 DNA tests. While they can be entertaining and educational, they have very little diagnostic value. Most of these tests only screen 1.9% of your genes, while we screen all your genes (or your entire exome, as scientists would say). Our competitor's Cystic fibrosis test, for example, only screens 33 markers. We test for 140 markers using our standard test, in addition to screening the entire gene associated with Cystic Fibrosis for any novel or known variant. We don't just offer this level of service for Cystic Fibrosis of course, but for more than 1700 other diseases included in our test.

Figure 2. Screenshot from <<https://www.gentlelabs.com/why-gentle>>, captured October 31, 2016. This company is no longer active, however archived versions of the page can be accessed by using the Wayback Machine at the Internet Archive <<https://archive.org/>>.

- The number of SNPs that would be analyzed in a research lab is much higher than the number of SNPs that commercial companies test for. For example, there are over 5000 SNPs that have been identified in the CYP2C9 gene, yet Gentle Labs only tests twelve of them and 23andMe tests two of them. Why would these companies only test a fraction of the known SNPs in a certain gene?

In 2013 the FDA sent a warning letter to 23andMe effectively banning the company from selling its genomic tests for health related traits. During this time, 23andMe in the United States could only sell its ancestry test kits. 23andMe in Canada, however, could sell both the ancestry kits and the health kits. Below are the links to the company's websites from the United States and Canada that detail the types of conditions that are tested for, as well as the warning letter issued by the FDA to 23andMe.

U.S. site: <<https://www.23andme.com/service/>> (US site)

Canadian site: <<https://www.23andme.com/en-ca/health/>> (Canadian site)

FDA warning letter: <<http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2013/ucm376296.htm>>

6. Read the FDA warning letter. What is the basis for the FDA ruling?

7. Do you agree with this ruling?

## Part II – Genomic Testing

Josie decided to have genomic testing through Gentle Labs. She received the kit and sent them a sample of her saliva. The company isolated DNA from the sample, and sequenced all the genes in her genome. Josie's DNA sequence was compared to a reference DNA sequence. Gentle Labs generated a report that categorized conditions based on their impact on metabolizing medication, health risks, lifestyle and carrier status.

### *Questions*

1. In the screenshot from the Gentle Labs website (Figure 2) what do they mean when they say “we sequence all your genes”? (*Hint: are all genes protein coding genes?*)

Below is some information on four of the traits Gentle Labs tests for: warfarin sensitivity, Alzheimer's disease, asparagus metabolite detection, and alcohol flush reaction.

### Warfarin Sensitivity

Warfarin is a blood thinning medication used to prevent the formation of blood clots and one of the most commonly used medications worldwide. Blood clots can block the flow of blood to important tissues and organs in the body, which can ultimately lead to tissue or organ damage. The dose of warfarin that an individual receives needs to be monitored accurately. Since warfarin is a blood-thinning medication, too much can cause excessive bleeding, while too little may not be effective in preventing blood clots. People vary in their sensitivity to warfarin—for those who are warfarin sensitive, a lower dose of medication is needed than for those who are warfarin resistant.

Variants in two genes, CYP2C9 and VKORC1, have been shown to affect individuals' sensitivity to warfarin. However, this sensitivity varies by ethnic group. Caucasians tend to be more sensitive to warfarin than Africans or Asians. For European Americans, about 30–40% of the variation is due to the genetic variants in CYP2C9 and VKORC1.

### Questions

- If 30–40% of the variation in warfarin sensitivity among European Americans is due to the genetic variants in CYP2C9 and VKORC1, then what is the remaining variation due to?
- How do you explain the difference in sensitivities to warfarin among different ethnic groups?

### CYP2C9

CYP2C9 encodes for a P450 2C9 isoenzyme, which normally acts to inhibit warfarin's anti-clotting properties. Isoenzymes are enzymes which differ in amino acid sequence, but catalyze the same reaction. This gene is found on chromosome 10. Two variants, CYP2C9\*2 and CYP2C9\*3, cause this enzyme to not function optimally, which in turn influences how readily warfarin is metabolized. For someone with either of these variants, warfarin is metabolized much slower and stays in the body longer. Therefore, individuals with the CYP2C9\*2 or \*3 variants need less warfarin than individuals that do not have these variants, but have variant CYP2C9\*1.

The P2C9\*2 variant is found in exon 3. The normal nucleotide at this position is a C. The \*2 variant has a T at this same position and results in a 30–40% decrease in enzyme activity.

*Genomic sequence with C variant:* ATT GAG GAC **C**GT GTT CAA GAG

*Genomic sequence with T variant:* ATT GAG GAC **T**GT GTT CAA GAG

- Translate each variant.

*C variant:*

*T variant:*

5. What type of mutation is this?

6. Hypothesize why the T variant would result in a 30–40% decrease in enzyme activity.

The CYP2C9\*3 variant is found in exon 7. The normal (reference) nucleotide at this position is an A. The \*3 variant refers to the presence of a C instead of an A in this position. This change results in an 80–90% decrease in enzyme activity. Both of these variants decrease the body's ability to metabolize warfarin. If the body cannot metabolize warfarin as well, warfarin will stay in the body longer. Therefore, people with these variants need lower doses of warfarin.

7. Another variant is found in the middle of the first intron of the CYP2C9 gene. The reference value is a T and the variant is a C variant. Would this variant change an amino acid? Why?

8. Of the three variants mentioned above, 23andMe only tests for the CYP2C9\*2 and \*3 variants, but not the variant in the intron. Why?

Interestingly, as a member of the cytochrome P450 gene family, CYP2C9 is involved in metabolizing a large number of medications such as NSAIDs (aspirin, ibuprofen, naproxen), some seizure medications, and some diabetic medications.

#### VKORC1

The VKORC1 gene, found on chromosome 16, codes for the enzyme vitamin K epoxide reductase, which helps convert vitamin K into a form that activates the clotting factors in the blood. Warfarin works by blocking this enzyme. If the enzyme cannot activate vitamin K, then clotting factors cannot be made. One variant in this gene, -1639G>A, also known as rs9923231, located in the promoter region of VKORC1 has been correlated with warfarin sensitivity. For this locus, the normal nucleotide is a G. The variant that reduces the amount of enzyme that is made is an A. Individuals with the A variant need less warfarin than individuals with the G variant. Each A variant increases a person's sensitivity to warfarin.

Proper dosing of warfarin depends on knowing an individual's genetic status at these different loci. Aside from the three different SNPs mentioned here, there are other SNPs in these two genes that affect the action of warfarin in the body. Different companies will test for a different number of these SNPs. For example, Gentle Labs tests for 13 different SNPs within the CYP2C9 and VKORC1 genes while 23andMe tests for CYP2C9\*2, \*3, and the VKORC1 SNP mentioned above. However, knowledge of one's genetic status alone is not enough. Individuals on warfarin therapy still need regular blood tests to monitor their clotting ability.

9. What other factors may influence an individual's ability to clot properly?

### *Late-Onset Alzheimer's Disease*

Alzheimer's disease is the most common form of dementia worldwide. Alzheimer's disease currently affects an estimated 2.5 to 4 million Americans, and this number is expected to increase in the coming decades as more people are living longer. In general, individuals are estimated to have a 10% lifetime risk of developing dementia, with 60% of dementia cases being caused by Alzheimer's disease. Alzheimer's disease starts with subtle memory loss, increasing forgetfulness or mild confusion as the only symptoms. However, over time, memory loss progressively worsens to a point where it interferes with most aspects of daily life.

Alzheimer's disease can be classified as early-onset or late-onset, depending on whether symptoms appear before or after age 65, respectively. Early-onset Alzheimer's disease accounts for less than 5% of all cases and results from clearly inherited mutations. Conversely, the development of late-onset Alzheimer's disease is thought to result from a combination of genetic, lifestyle and environmental factors.

10. Late onset Alzheimer's disease is considered a multi-factorial trait. What are the other factors (apart from genetic causes) that contribute to an individual's overall risk?

The exact cause of late-onset Alzheimer's disease is currently unknown. This form of the disorder probably results from a combination of genetic and environmental factors. A variant of the apolipoprotein E gene, APOE, has been recognized and extensively studied as a genetic risk factor for the development of late-onset Alzheimer's disease. Three common variants of the APOE gene exist in the general population, called APOE  $\epsilon$ 2,  $\epsilon$ 3 and  $\epsilon$ 4. APOE  $\epsilon$ 3 is the most frequent variant in the general population. APOE  $\epsilon$ 4 is common in Northern Europe, whereas APOE  $\epsilon$ 2 is a rare variant. The APOE  $\epsilon$ 4 variant has been consistently associated with an increased risk of developing late-onset Alzheimer's disease. More specifically, carrying a single copy of the APOE  $\epsilon$ 4 allele increases the risk approximately threefold, while carrying two copies of this allele increases the lifetime risk more than tenfold. On the other hand, carrying the APOE  $\epsilon$ 2 allele may have a small protective effect against developing late-onset Alzheimer's disease in certain populations. How different APOE alleles exactly impact the risk of developing late-onset Alzheimer's disease is still unclear.

The APOE gene encodes a protein that is a major constituent of lipoproteins, which are responsible for packaging and transporting cholesterol and other fats through the bloodstream and cerebrospinal fluid. Defects in lipid and cholesterol trafficking, or a defect in the production, aggregation or clearance of amyloid beta plaques are mechanisms proposed to contribute to the role of the APOE protein in Alzheimer risk.

It is important to note that the presence of an APOE  $\epsilon$ 4 allele only changes the risk of developing late-onset Alzheimer's disease. Not all carriers of one or two APOE  $\epsilon$ 4 alleles will develop Alzheimer's disease, and not having any APOE  $\epsilon$ 4 still confers a general population lifetime risk of developing late-onset Alzheimer's disease. This is due to other, yet unknown risk factors.

11. Based on the function of the APOE protein, do you think there are any other diseases that may be associated with defects in APOE?



APOE is located on chromosome 19. Two SNPs, rs429358 and rs7412, are used to determine the three APOE variants,  $\epsilon 2$ ,  $\epsilon 3$ , and  $\epsilon 4$ . Below are two examples of how the SNPs are used to determine APOE variants. The fourth variation is extremely rare and therefore not included.

*Example 1:* An individual that has the CT genotype for both SNPs is almost always  $\epsilon 2/\epsilon 4$ .

Variant	rs429358	rs7412
$\epsilon 2$	T	T
$\epsilon 3$	T	C
$\epsilon 4$	C	C

*Example 2:* An individual that has the CT genotype at rs429358 and CC genotype at rs7412 is almost always  $\epsilon 3/\epsilon 4$ .

Variant	rs429358	rs7412
$\epsilon 2$	T	T
$\epsilon 3$	T	C
$\epsilon 4$	C	C

### Asparagus Metabolite Detection

Some individuals are able to detect a strong odor in their urine after eating asparagus while other individuals are not. The odor is due to an excreted sulfur-containing metabolite called *methanethiol*. It has yet to be determined whether smelling the asparagus scent in urine is due to the ability to produce methanethiol or the ability to detect it. In any case, there seems to be a genetic component to this strange trait. There is a strong association with a genetic marker, rs4481887, where each copy of an A allele increases the risk of being able to smell asparagus in urine 1.67 times as compared to individuals with a GG at this marker.

### Alcohol Flush Reaction

Alcohol can have an immediate, unpleasant reaction in some individuals. Common signs of alcohol intolerance are skin flushing (redness), nasal congestion, headache, low blood pressure, nausea and vomiting. This condition is often referred to as alcohol flush reaction caused by a genetic condition in which the body cannot break down alcohol. The only way to prevent the reaction is avoidance. The reaction is due to variations in two genes that encode proteins for breaking down alcohol in the bloodstream.

Alcohol flush reaction is determined by genotype in the genes ALDH2 and ADH1B. ALDH2 codes for an enzyme called *aldehyde dehydrogenase*. This enzyme is involved in processing highly toxic acetaldehyde to a harmless acetic acid. Having an A allele at the SNP inactivates the enzyme and prevents conversion to acetaldehyde. If an individual has the AA genotype, the individual is highly sensitive to alcohol and the acetaldehyde is removed very slowly from the body. The other gene is not tested, however, alcohol sensitivity may be altered even greater depending on an individual's genotype at ADH1B.

- One possible treatment for alcoholics is the drug called disulfiram (Antabuse). Research the mechanism of action of disulfiram and explain why it would be a potential treatment. (*Suggested resource:* <<http://www.disulfiram.org/>>.)

## Part III – Lab Results

Below is a partial list of Josie’s results showing the gene that was tested, the chromosomal location of each gene and the genotype of the SNPs within that gene.

### Report

<i>Variant tested</i>	<i>Genotype</i>	<i>Chromosomal Location</i>	<i>Phenotype/Risk</i>
APOE (rs429358)	CT	19	
(rs7412)	CC	19	
VKOR1 (rs9923231)	AA	16	
CYP2C9*2 (rs1799853)	CC	10	
CYP2C9*3 (rs1057910)	AC	10	
Marker rs4481887	AA		
ALDH2 (rs671)	AG	12	

### Questions

- Using the information you learned about the various traits in Part II, fill in the final column of the chart above by determining Josie’s phenotype or risk based on her genotype. Explain how you interpreted the results.
  - What other factors will alter Josie’s risk of developing Alzheimer’s disease?
  - Based on Josie’s genotype, what dose of warfarin would be best for her (high, low, or intermediate dose)?
  - One of Josie’s favorite vegetables is asparagus. She noticed that she could detect a strong sulfur smell in her urine after eating asparagus. Explain why she is able to do this after looking at her data.
  - By looking at the data, what does Josie’s genotype tell you about her sensitivity to alcohol? Should this information result in any behavior modifications?
- Typically, genomic testing services group the traits they test for into categories such as (1) recessive disorders, (2) drug response, (3) risk factors, and (4) non-health related traits. Consider the four traits mentioned in Part II (Alzheimer’s disease, warfarin sensitivity, asparagus metabolite detection, and alcohol flush reaction). In which category does each trait belong?
- Would you have liked to know your APOE status regardless of the results? Why or why not?
- Do you think that Josie should be obligated to tell her family members the results? Why or why not? Would you feel differently based on different APOE results? Would you feel differently if you were a carrier for a recessive disease? If you had the allele for Huntington’s disease, an autosomal dominant disorder characterized by progressive neurodegeneration?

## Part IV – Conclusion

After receiving her report from Gentle Labs, one of the genetic counselors on staff at Gentle talked with Josie at length about her results and what they meant in terms of her overall health and any preventative measures she could take in order to stay healthy. Josie decided to visit her parents to discuss her results with them. Her paternal grandmother, who just turned 80 years old, was diagnosed with the beginning stages of Alzheimer’s disease. Josie wanted to encourage her father and his siblings to get tested to see if they were at an increased risk of developing Alzheimer’s disease as well. She wanted to know what her father would want her to do if she saw any early signs of Alzheimer’s disease.

Together, Josie and her father decided that they would inform their family physician about Josie’s test results. Although Josie was a healthy young adult, the report from Gentle Labs presented her with actionable results: her sensitivity to warfarin could directly impact medical care she might receive in the future. Her sensitivity to alcohol might or might not impact her social behavior.

### Question

1. If you were Josie, how would this information impact your lifestyle choices or healthcare planning?



### References

- Bagyinszky, E., Y.C. Youn, *et al.* 2014. The genetics of Alzheimer’s disease. *Clin Interv Aging* 9: 535–551. doi: 10.2147/CIA.S51571
- Bush, W.S. and J.H. Moore. 2012. Genome-wide association studies. *PLoS Comput Biol.* 8: e1002822.
- Crabb, D.W., H.J. Edenberg, *et al.* 1989. Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH2(2) allele is dominant. *J Clin Invest.* 83:314–316.
- Jorgensen, A.L., R.J. FitzGerald, *et al.* 2012. Influence of CYP2C9 and VKORC1 on patient response to warfarin: a systematic review and meta-analysis. *PLoS One* 7(8): e44064. doi: 10.1371/journal.pone.0044064.
- Richards, C.S., S. Bale, D.B. Bellissimo, S. Das, W.W. Grody, M.R. Hedge, E. Lyon, and B.E. Ward 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet. Med.* 10, 294–300.