Genomic Medicine and Polygenic Risk Scores



## Part 1

In Part 1, we’ll go over the background and data for this activity.

Many human diseases are caused by a combination of genetic and environmental factors. For most of these diseases, it can be difficult to predict what combination will lead to a person developing disease. Instead, physicians and researchers talk about the increased risk - essentially, we know what factors increase the likelihood of developing diseases, even if we can’t predict with 100% certainty whether someone might get sick.

### Polygenic Inheritance

When we think about genetically-inherited diseases, we usually think about those where a single gene variant can cause disease. Most of the classic examples of genetic disease, like sickle cell anemia, Tay Sachs, or Huntington’s, are passed along via *Mendelian inheritance*. A mutation in a single gene is enough to disrupt normal protein synthesis and cellular processes, causing a person to become sick.

However, most diseases, particularly common diseases, are not the result of a mutation to a single gene. Instead, they are the result of mutations to many different genes. Each mutation itself isn’t enough to cause problems, but when all of them show up together, the cellular processes aren’t able to function normally. This type of inheritance (when a trait is caused by a combination of mutations to many genes) is called *polygenic inheritance*. Polygenic traits are quite common, and they aren’t always diseases! Any trait that shows a range of phenotypes is likely to be the result of polygenic inheritance. Some of the more famous polygenic traits include height and skin color.

### Polygenic Risk Scores

Researchers have developed something called a “polygenic risk score” (PRS) to estimate how likely a person might be to develop a disease based on their genetics.

|  |
| --- |
| Tip |
| It’s always important to remember that **having a higher PRS does not guarantee someone will get a disease.** Likewise, having a lower PRS does not guarantee someone is protected from a disease.  A PRS is just a rough estimate based on what we know at the moment about the genetics that contribute to developing a disease. In many case, a high PRS will only increase a person’s overall risk for a disease by less than 5%. |

At its core, the polygenic risk score is a simple idea. If multiple variants can contribute to the development of a disease, other variants might also provide protection from a disease. We can look at all the variants a person might have, add them up, and then get an idea of whether they have more variants that contribute to a disease relative to how many protective variants they have.

Being able to calculate a PRS requires that scientists have an understanding of all the genes and noncoding regions of DNA that might contribute to developing (or not developing) a disease. **A PRS is only as good as the reference database for a particular disease or trait.**

### Genomic Ancestry

### Exploring Variant Data

In the next steps, we’ll be looking at how PRS can help patients be more informed about disease risk.

This example will look at prostate carcinoma, or prostate cancer. While prostate cancer is common in men and is a leading cause of cancer-related death, it tends to be slow growing with limited aggressiveness (see <https://www.ncbi.nlm.nih.gov/books/NBK470550/>). This means that genetic screening and symptom monitoring, especially in older age, can have a big impact on outcomes.

### Mr. J’s Data

We’re going to take gene screening data from an imaginary patient, Mr. J, to understand his risk. Mr. J has African ancestry, which will be important later.

Get the data at <https://genomicseducation.org/data/prs_ind_1.txt>. It should look like:

rs7463326:G  
rs58235267:G  
rs74001374:C  
...

|  |
| --- |
| What does the data mean? |
| Each line of the data represents a variant location in the genome and individual person’s version of that gene (allele).  For example, **rs7463326** is the variant. Mr. J has a “G” here. Other patients might have “A” or “C”. |

### Exploring Variants

Let’s explore the **rs7463326** variant a bit more. Go to <https://www.ebi.ac.uk/gwas/variants> and type “rs7463326” in the search bar.



Select the first result.



Notice the variant information. The most severe **rs7463326** variant lies in an [intergenic region](https://www.genome.gov/genetics-glossary/Intergenic-Regions#:~:text=Definition,Medical%20Genetics%20Branch), or between protein coding genes.



Scroll down to see the risk allele associated with the variant. For **rs7463326**, the risk allele is “G” and it’s associated with prostate carcinoma.



Mr. J has a “G” here. Other patients might have “A” or “C”. This means he might be more at risk than other patients. However, we know risk is often dictated by multiple genes. In the next steps we’ll explore how to assess risk with multiple variants.

|  |
| --- |
| Check Your Knowledge |
| 1. Try looking up another variant at <https://www.ebi.ac.uk/gwas/variants>. Look up “rs2075650”. What kind of variant is listed under “Most severe consequence”? Are there any diseases associated with this variant? |

## Part 2

In Part 2, we’ll calculate our first PRS score.

### Calculating Risk

Go to https://prs.byu.edu/calculate\_score.html - this is the website we’ll use to calculate a PRS for Mr. J’s data.



Return to the data at <https://genomicseducation.org/data/prs_ind_1.txt>. Copy this data into the top part of the PRS calculator website.