When genome is copied to make a new cell it makes mistakes and causes small variations in the DNA sequence, this is called a (Single Nucleotide Polymorphism) SNP. There are about 10million possible SNP’s in the human genome which account for our differences. These are categorized by rfID’s in 23andme’s raw data. The rfid’s do not tell us where the variation is located, but only the specific variation. They can tell us how we look, respond to diseases, and respond to specific drugs, etc. However, most SNP’s hold no observable differences. The rfID data is all stored on ncbi’s(national center for biotechnology information) data base which anyone can add to. Many SNP’s are identified, but not evaluated for significance. Most significance is based on papers that notice some sort of correlation. Anyone can contribute.

Example Report

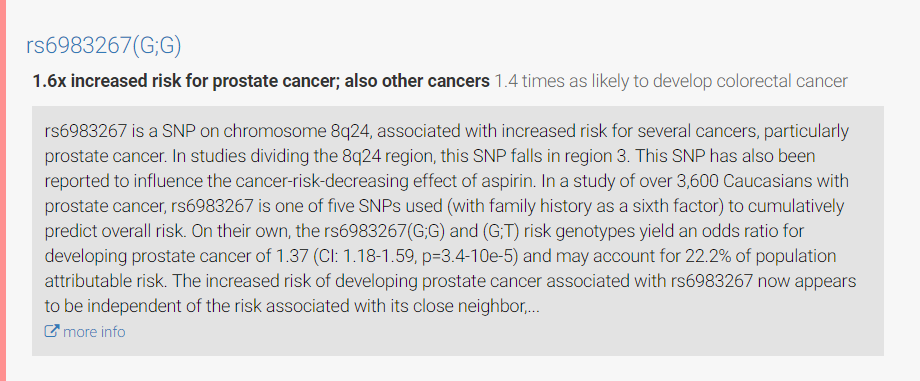
<http://files.snpedia.com/reports/promethease_data/genome_Lilly_Mendel_v4_ui2.html>

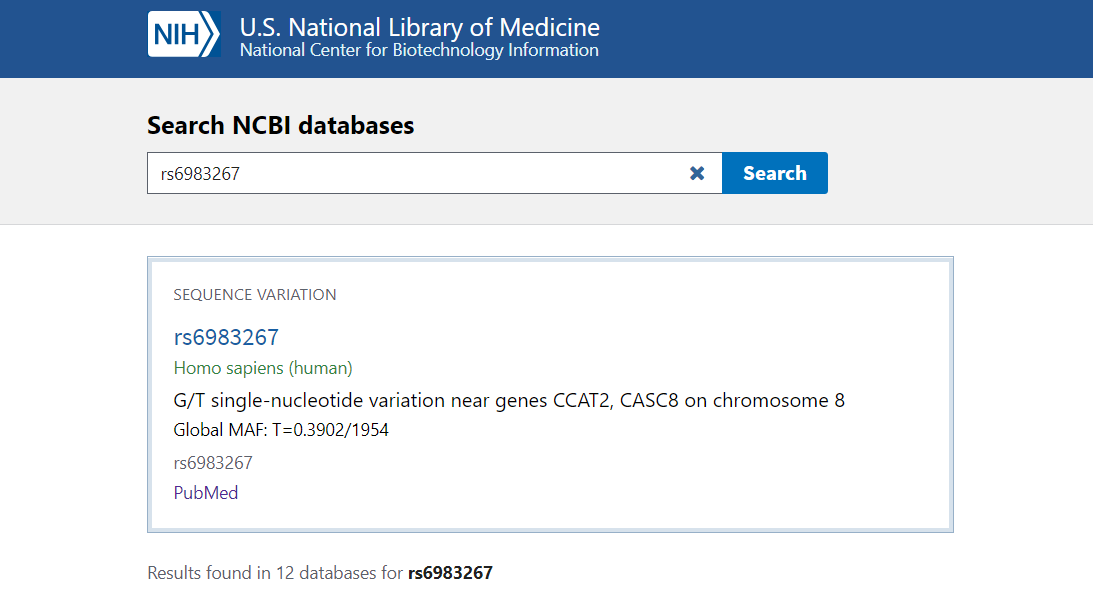
-Most likely automate the parsing of data process below.

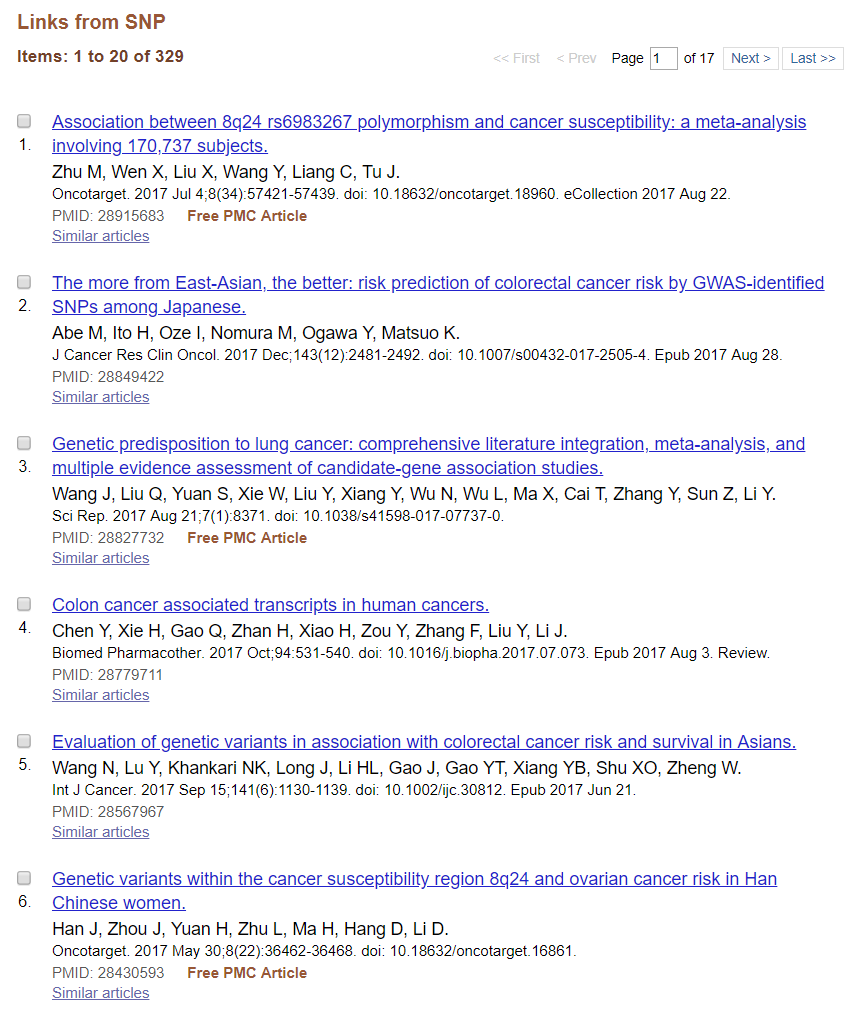
The data provided in the report is based on 2015

-Some rsIDs are no longer associated with the same meaning as in 2015.

-One of the more clear examples from Promethease:







Explains what rfID’s are

<https://www.ncbi.nlm.nih.gov/books/NBK44417/#Content.what_is_a_reference_snp_or__rs_i>

-Finds

Database of rfid’s

<https://www.ncbi.nlm.nih.gov/snp/>

Biopython-Used for computational molecular biology. Can be used to fetch rfID data.

<https://github.com/biopython/biopython>

<http://biopython.org/DIST/docs/tutorial/Tutorial.html>

<https://en.wikipedia.org/wiki/BLAST>

In bioinformatics, BLAST for Basic Local Alignment Search Tool is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.