Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Subject Section

Nutritional Genomics Visualization Tool on 23andMe Raw Genotyping Data

Natalie Wang[1,*], Celine Hoh[2] and Michael Schatz[2]

[1]Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, [2]Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, [3]Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD.

Abstract

Motivation: While the genomic testing technology 23andMe is becoming more prevalent, very few personalized dietary guidelines based on genetic characteristics are present in the market yet.
Results: We created a database and visualizations on Single Nucleotide Polymorphisms (SNPs) that are related to taste and smell preferences, as well as sensitivities and metabolic differences.

Contact: choh1@jhu.edu, nwang42@jhu.edu
Supplementary information: Supplementary data would be provided in the final project paper.

1 Introduction

Many past and current dietary guidelines follow a one-size-fits-all approach, where the effect of an individual's genomics on nutrient metabolism is not taken into account. However, with genomic technologies like 23andMe becoming increasingly prevalent today, we aim to use this genetic testing tool to learn more about an individual's food preference and nutrient absorption using their genomics markers with known associations.

The 23andMe results provide us with a list of Single Nucleotide Polymorphisms (SNPs) that are present in an individual. SNP is a site in the DNA sequence where an individual's genome differs by one base pair from the reference genome[1]. Humans are generally 99.8-99.9% similar to each other, which results in about a 1 millions differences in base pairs. The Genome-Wide Association Study (GWAS) is an observational study of a genome wide set of genetic variants and their association to a trait[2]. Hence, by checking for an individual's SNPs from GWAS, we aim to use this genetic variation to identify food preferences and nutrients susceptibility.

Nutritional genomics is mainly divided into two main categories, namely nutrigenomics and nutrigenetics. Nutrigenetics studies the effect of genetic variation on dietary response and nutrient requirements, while nutrigenomics considers how food components affect gene expression and subsequent proteomic/metabolomic changes[3]. In this study, we mainly focused on the nutrigenetics field. We would look into an individual's taste and smell preference, as well as their sensitivities and metabolism rates towards different foods and nutrients by analyzing their SNPs from 23andMe results. We would then provide personalized recipe recommendations based on the individual's food preference and metabolism/nutrient absorption rate.

2 Methods

This project seeks to help users visualize and understand their 23andMe results specifically around nutrition and health. The main areas we explored were taste/food preferences, smell preferences, food sensitivities and metabolic differences.

2.1.1 23andMe Datasets

The raw genomic datasets used to build and test our visualizations are from OpenSNP[4], an open platform for collecting commercial genomic data aimed at crowd-sourced association studies to create new knowledge about our genes. We selected participants from this pool based on the recency of their genomic data uploads (we wanted to ensure we had users with the most recent 23andMe data format) as well as ensuring they had filled out the optional phenotype survey which includes questions about food preferences. We selected these users so we would be able to validate and sanity check our reports.

The current version of 23andMe raw data includes the rsid of found SNPs, the chromosome it is located on, the position it is found at, and finally the genotype. We are able to use this data to match these SNPs to associations with various nutrition-related factors such as taste preferences.

2.1.2 SNP Lookups

We compiled a database of nutrition-specific SNP associations with a literature review, SNPedia[5], and DbSNP[6]. SNPedia is a wiki that shares information about the effects of variations in DNA and provides links to peer-reviewed scientific articles. DbSNP is the largest public database of SNPs and is maintained by the NIH. Our database includes the RSID of the SNPs, gene if affects, type of outcome, the outcome, and the link to the source. We intend to make this database public on Kaggle once we finish this project.

2.2 Analysis/Visualizations

This project focuses on providing new analysis and visualizations for the users. We aim to make clear, straightforward visualizations that are easy to read and understand. While the presence of an SNP does not mean a person will definitely express a certain phenotype, it does increase the likelihood of its expression so we chose to analyze SNPs found by 23andMe.

2.2.1 Overview

The first set of visualizations we will provide for the user will be summary visualizations about their specific genomic data. This will include a barchart of the number of SNPs per chromosome as well as a brief explanation of each chromosome. This section will also have an overview Circos chart so users will be able to visualize their whole genome.

2.2.2 Nutrition and Health

Taste is fundamental to health because it can have a large impact on food preferences - what tastes people prefer inform their food choices. For the first part of our analysis, we look at each user's SNPs that are located on taste receptor genes. Some of the studied taste preferences include preferences or aversions for sweet, salty, umami, sour, and bitter[7]. Additionally we included information from studies that found associations between specific SNPs and food preferences such as opinions on artichokes, broccoli, bacon, ice cream, coffee, etc[8]. We will create charts indicating possible preferences users may have.

We also looked at smell preferences as this could also have an impact on food preferences and health habits around eating. Some of the study results we included looked at preferences for certain foods like pork, licorice, cinnamon, lemongrass, and oranges[9]. We will also create visualizations for users indicating what smell preferences they may be more likely to have. We also use this information later to help inform our recipe recommender.

Another important aspect of nutrition is food sensitivities and metabolic differences. Using several studies on the effects of specific SNPs on food sensitivities, we are able to create a list of SNPs associated with various food sensitivities like caffeine sensitivity and alcohol sensitivity. We were also able to find SNPs related to various nutrition deficits such as iron deficiencies, folic acid deficiencies, and vitamin C deficiencies[10]. We will create a visualization based on this information to show users their SNPs and what they are associated with so they might make more informed choices.

3 Results

The data being evaluated is an individual's genomic data uploaded to OpenSNP. We chose this individual as our first test case and to help develop the visualizations based on recency of upload and because she filled out the optional phenotype survey. For preliminary results, we have started creating the summary visualizations we will include with each report (Fig. 1 below). We are in the process of finishing our SNP database and will begin creating the visualizations for taste/food preferences, smell preferences, and sensitivities and metabolic differences once that is complete.

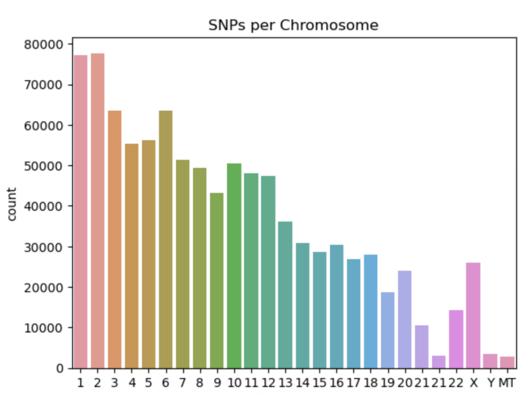

Fig. 1. Number of SNPs per Chromosome. First 23andMe sample.

4 Discussion

By creating an extensive database and visualizations on nutrigenetics related SNPs, we could help individuals navigate their journey towards achieving their optimal health while taking into account their preferences. Besides benefiting individuals who are seeking personalized dietary recommendations, over time our analysis would also help improve public health guidelines by linking scientific evidence on diet to health.

Nevertheless, there are two factors that we need to take into consideration for this study. Firstly, while genomic variation has a promising future in giving dietary recommendations based on an individual's genetic characteristics, there are many other physiological and environmental factors that influence an individual's preferences and metabolism rate too[11]. Hence, we aim to also integrate personal characteristics such as gender, physiological state, physical activity, allergies, and other special conditions alongside genetic variants to provide more holistic dietary advice. It is also important to note that nutritional health is heavily dependent on other environmental factors such as availability and coingestion of other nutrients, and such interactions could lead to different requirements for individuals.

Secondly, nutritional genomics is a new and evolving field that still needs much work on. While GWAS, the library we reference our SNPs from, is a well established international consortium for discovering genetic variants that contribute to complicated diseases, many nutrient and diet related variants are still missing. Thus, as more nutrient-SNP associations are being discovered, the development of dietary advice would also become increasingly accurate.

Acknowledgements

We would like to thank Prof. Michael Schatz for providing us the opportunity to work on this project.

Conflict of Interest: none declared.

References

1. Gill, Peter. "An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes." International journal of legal medicine 114.4 (2001): 204-210.
2. Uffelmann, E., Huang, Q.Q., Munung, N.S. et al. Genome-wide association studies. Nat Rev Methods Primers 1, 59 (2021). https://doi.org/10.1038/s43586-021-00056-9
3. DeBoer, Anna. "Nutrigenomics: Moving towards Personalized Nutrition for Obesity Prevention." Health Science Inquiry 3.1 (2012): 16-Page.
4. Greshake, Bastian, et al. "OpenSNP–a crowdsourced web resource for personal genomics." PloS one 9.3 (2014): e89204.
5. Cariaso, Michael, and Greg Lennon. "SNPedia: a wiki supporting personal genome annotation, interpretation and analysis." Nucleic acids research 40.D1 (2012): D1308-D1312.
6. Sherry, Stephen T., et al. "dbSNP: the NCBI database of genetic variation." Nucleic acids research 29.1 (2001): 308-311.
7. Chamoun, Elie et al. "The Relationship between Single Nucleotide Polymorphisms in Taste Receptor Genes, Taste Function and Dietary Intake in Preschool-Aged Children and Adults in the Guelph Family Health Study." Nutrients vol. 10,8 990. 29 Jul. 2018, doi:10.3390/nu10080990
8. Robino, Antonietta et al. "A Brief Review of Genetic Approaches to the Study of Food Preferences: Current Knowledge and Future Directions." Nutrients vol. 11,8 1735. 26 Jul. 2019, doi:10.3390/nu11081735
9. Li, Bingjie, et al. "From musk to body odor: Decoding olfaction through genetic variation." PLoS genetics 18.2 (2022): e1009564.
10. Delgado-Lista, Javier, et al. "Top single nucleotide polymorphisms affecting carbohydrate metabolism in metabolic syndrome: from the LIPGENE study." The Journal of Clinical Endocrinology & Metabolism 99.2 (2014): E384-E389.
11. Mathers, John C. "Nutrigenomics in the modern era." Proceedings of the Nutrition Society 76.3 (2017): 265-275.