h+: the future of transgenic technologies is here

Vinogradova Sofiya, Zherko Nikita
Moscow Institute of Physics and Technology

February 17, 2024

Abstract

Analysis of raw 23&Me data unveils key genetic insights into ancestry and disease predispositions. In this study, the individual was identified as belonging to the H(T152C) mtDNA haplogroup and the R1a1a Y haplogroup, common in Europe. Additionally, personal characteristics of male sex and brown eyes were confirmed. Of particular interest was the identification of rs231775, associated with CTLA-4 gene, associated with a 1.5-fold increase in Graves' disease risk and linked to diabetes mellitus. CRISPR-based A-to-G editing is proposed to mitigate these risks. Furthermore, the presence of Rs4961 within ADD1 gene, associated with increased blood pressure risk, prompted consideration for base editing to reduce susceptibility. Overall, this study underscores the potential of genetic analysis for personalized interventions, such as CRISPR-based editing, to mitigate disease risks and improve individual health outcomes in future.

1 Introduction

Single nucleotide polymorphisms (SNPs) represent the most common type of genetic variation among humans. In an average genome, there are about 4 to 5 million variances compared to the standard human genome, with the vast majority (over 99.9%) being SNPs and indels [1].

SNPs located within genes or regulatory regions adjacent to genes can have a direct impact on disease by altering gene function. Overall, SNPs serve as biological markers, assisting in predicting an individual's response to certain medications, susceptibility to environmental factors, risk of developing diseases, family medical history, and inheritance patterns [2, 3].

The primary method for efficiently identifying known SNPs is through the utilization of microarrays, while Next-Generation Sequencing (NGS) is employed for the discovery of new SNPs and variants. Microarray-based genotyping for SNPs entails the use of known nucleotide sequences as probes to hybridize with tested DNA sequences, allowing for qualitative and quantitative SNP analysis through

signal detection [4]. These arrays can be tailored to specific research goals and offer numerous advantages, including high throughput, simultaneous processing of multiple samples, and the ability to assess known markers within the human genome. These features make SNP microarrays well-suited for large-scale population studies and Genome-Wide Association Studies (GWAS) [3].

When a disease-associated SNP is identified, it can potentially targeted using CRISPR-Cas editors for therapeutic purposes, particularly in the context of allele-specific treatment of autosomal dominant disorders. CRISPR-Cas editors can be used to selectively target and correct the mutated allele while leaving the wild-type allele intact. First approach addresses SNPs that generate or eliminate PAM sites, CRISPR-Cas system can be designed to recognize and edit those [5, 6]. Another approach to target SNPs is usage of CRISPR-Cas base editors, which can directly convert one nucleotide to another without inducing double-stranded DNA breaks. This approach has been demonstrated in preclinical studies targeting specific mutations in diseases such as sickle cell anemia [7].

2 Materials and Methods

The raw 23&Me data was downloaded from [8]. This data was prepossessed to save only SNPs with the plink tool (version 1.9) [9, 10]. The following parameters were used: –23file –recode vcf –output-chr MT –snps-only just-acgt. The maternal mtDNA haplogroup was revealed by the James Lick online tool [11] and the parental Y haplogroup by the Y-SNP Subclade Predictor [12]. Eye color annotation was performed using the HIrisPlex-S Eye, Hair and Skin Colour DNA Phenotyping Webtool [13, 14]. Annotation for all SNPs was performed using the Variant Effect Predictor online tool [15, 16] for the 37 genome assembly. Additional information on the annotated SNPs was searched on the SNPedia website [17].

3 Results

The 23&Me data was subjected to mtDNA and Y DNA haplogroup analysis. The mtDNA analysis revealed belonging to the following groups: H(T152C), H1(T152C), H, H16(T152C), H3(T152C), H46, H52, H69 or H9. The Y DNA analysis revealed belonging to the R1a1a, M3, K, R1b1a2a1a2a1b3$\sim$ 2 or N1a groups.

Based on the rs12913832, rs1800407, rs12896399, rs16891982, rs1393350, rs12203592 SNPs the brown color of eye with the p-value of 0.012 was revealed using the online tool. The sex was determined as male.

The annotations of all SNPs using the VEP online tool are present in the lab journal. The most interesting ones are shown in the Table 1 below.

SNP ID	Genotype	Clinical significance
rs1049296	C/T	Heterozygote carrying both C1 and C2 transferrin subtypes; very slightly higher risk for Alzheimers
rs12150220	A/T	slightly increased risk for several autoimmune diseases
rs13266634	C/T	increased risk for type-2 diabetes
rs1801274	A/G	complex; generally greater risk for cancer progression
rs2241880	A/G	1.4x increased risk for Crohn's disease in Caucasians
rs231775	A/G	1.5x risk of autoimmune thyroiditis
rs4961	G/T	1.8x increased risk for high blood pressure
rs5174	C/T	1.3x increased risk for heart disease
rs6265	C/T	Slightly increased risk for ADHD or depression; somewhat quicker mental decline in Alzheimer patients
rs699	A/G	increased risk of hypertension

Table 1: Top-10 interesting SNPs and their clinical significance

4 Discussion

The genetic analysis results obtained from the raw 23&Me data provide valuable insights into the individual's ancestry and predispositions to certain diseases.

The individual more likely belongs to the H(T152C) mtDNA haplogroup, one of the most common mtDNA clade in Europe. The identified R1a1a Y haplogroup is also one of the most common in Europe and probably originates from the region within the Eurasian Steppes or the Middle East [18]. The sex and eye color of the individual was determined as male and brown, respectively. Considering that we personally know this individual, the reveled features are likely to be true.

Among the identified mutations, we considered rs231775 the most interesting and relevant one. The CTLA-4 rs231775 G/G genotype is associated with a 1.5-fold increase in the risk of Graves' disease, a form of autoimmune thyroiditis [19]. It has also been found that this mutation contributes to the occurrence of diabetes mellitus [20]. So, we propose editing A-to-G CRISPR base editing to get a gene that has no risk of Graves' disease.

Another interesting SNP is Rs4961 located in the ADD1 gene that promotes the conversion of glycine to tryptophan. This mutation causes increased blood pressure in the presence of the T allele (risk carrier) [21]. In our case, we are dealing with the G/T form, which increases thw risk for high blood pressure by 1.8-fold. However, it is possible to perform base editing to achieve the innocuous G/G genotype.

5 References

[1] Adam Auton et al. "A global reference for human genetic variation". In: *Nature* 526 (7571 Oct. 2015), pp. 68–74. ISSN: 0028-0836. DOI: 10.1038/nature15393.

[2] Emil Uffelmann et al. "Genome-wide association studies". In: *Nature Reviews Methods Primers* 1 (1 Aug. 2021), p. 59. ISSN: 2662-8449. DOI: 10.1038/s43586-021-00056-9.

[3] GWAS. URL: www.ebi.ac.uk/gwas/.

[4] DNA microarrays. URL: https://www.cd-genomics.com/the-principles-and-workflow-of-snp-microarray.html.

[5] D G Courtney et al. "CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting". In: *Gene Therapy* 23 (1 Jan. 2016), pp. 108–112. ISSN: 0969-7128. DOI: 10.1038/gt.2015.82.

[6] Jun Wan Shin et al. "PAM-altering SNP-based allele-specific CRISPR-Cas9 therapeutic strategies for Huntington's disease". In: *Molecular Therapy - Methods Clinical Development* 26 (Sept. 2022), pp. 547–561. ISSN: 23290501. DOI: 10.1016/j.omtm.2022.08.005.

[7] Gregory A. Newby et al. "Base editing of haematopoietic stem cells rescues sickle cell disease in mice". In: *Nature* 595 (7866 July 2021), pp. 295–302. ISSN: 0028-0836. DOI: 10.1038/s41586-021-03609-w.

[8] 23&Me raw data. URL: https://drive.google.com/file/d/1QJkwJe5Xl_jSVpqdTSNXP7sqlYfI666j/view.

[9] PLINK 1.9. URL: www.cog-genomics.org/plink/1.9/.

[10] Christopher C Chang et al. "Second-generation PLINK: rising to the challenge of larger and richer datasets". In: *GigaScience* 4.1 (Feb. 2015), s13742-015-0047–8. ISSN: 2047-217X. DOI: 10.1186/s13742-015-0047-8. eprint: https://academic.oup.com/gigascience/article-pdf/4/1/s13742-015-0047-8/25512027/13742_2015_article_47.pdf. URL: https://doi.org/10.1186/s13742-015-0047-8.

[11] James Lick online tool. URL: https://dna.jameslick.com/mthap/.

[12] Y-SNP Subclade Predictor. URL: https://ytree.morleydna.com/#.

[13] Hair HIrisPlex-S Eye and Skin Colour DNA Phenotyping Webtool. URL: https://hirisplex.erasmusmc.nl.

[14] Lakshmi Chaitanya et al. "The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation". In: *Forensic Science International: Genetics* 35 (July 2018), pp. 123–135. ISSN: 18724973. DOI: 10.1016/j.fsigen.2018.04.004.

[15] Variant Effect Predictor. URL: http://grch37.ensembl.org/Homo_sapiens/Tools/VEP.

[16] William McLaren et al. "The Ensembl Variant Effect Predictor". In: *Genome Biology* 17 (1 Dec. 2016), p. 122. ISSN: 1474-760X. DOI: 10.1186/s13059-016-0974-4.

[17] SNPedia. URL: https://www.snpedia.com.

[18] Horolma Pamjav et al. "Brief communication: New Y-chromosome binary markers improve phylogenetic resolution within haplogroup R1a1". In: *American Journal of Physical Anthropology* 149 (4 Dec. 2012), pp. 611–615. ISSN: 0002-9483. DOI: 10.1002/ajpa.22167.

[19] Edyta Pawlak-Adamska et al. "CD28/CTLA-4/ICOS haplotypes confers susceptibility to Graves' disease and modulates clinical phenotype of disease". In: *Endocrine* 55 (1 Jan. 2017), pp. 186–199. ISSN: 1355-008X. DOI: 10.1007/s12020-016-1096-1.

[20] D D Abramov et al. "Polimorfizm gena CTLA4 (49A/G) v russkoy populyatsii u bol'nykh sakharnym diabetom 1 tipa i zdorovykh donorov". In: *Diabetes mellitus* 10 (3 Sept. 2007), pp. 2–3. ISSN: 2072-0378. DOI: 10.14341/2072-0351-5989.

[21] Daniele Cusi et al. "Polymorphisms of -adducin and salt sensitivity in patients with essential hypertension". In: *The Lancet* 349 (9062 May 1997), pp. 1353–1357. ISSN: 01406736. DOI: 10.1016/S0140-6736(97)01029-5.