Exploring Human Ancestry, Phenotypic Traits, and Clinical Relevance Through Genome SNP Analysis

M. Uzun[a,c], A. Fedorenko[b,c]

[a]*Research Center Biotechnology RAS, Moscow, Russia*
[b]*Skolkovo Institute of Science and Technology, Moscow, Russia*
[c]*Bioinformatics Institute, Saint Petersburg, Russia*

Abstract

Advancements in genomic technologies, including the Human Genome Project and high-throughput sequencing, have transformed human genome studies, unveiling insights into genetic structure and heredity. Genome-wide association studies have emerged as powerful tools for understanding complex diseases by identifying genetic variants. The revolutionary CRISPR/Cas9 technology offers precise genome editing, promising to address inherited disorders. However, ethical debates surround its use for genetic enhancement. This review analyzes a young woman's genome, uncovering ancestry through haplogroup analysis and exploring single nucleotide polymorphisms related to phenotypic traits and potential health implications. Findings illuminate the intricate relationship between genetics and observable characteristics.

Keywords: Single nucleotide polymorphisms, Human genome, Haplogroup analysis, Genome editing

1. Introduction

The completion of the Human Genome Project and HapMap, along with high-throughput sequencing, broaden the opportunities in human genome studies. Since the landmark achievement of sequencing the first entire human genome in 2003, researchers have not only elucidated the structure and organization of the three billion base pairs of DNA but have also pioneered novel approaches to investigate the hereditary nature of humanity [1]. Genome-wide association studies (GWAS) have emerged as powerful tools for understanding the genetic underpinnings of complex diseases such as cancer, cardiovascular diseases, and neurological disorders by identifying single-nucleotide polymorphisms (SNPs) associated with these conditions [2]. This integration of genomic data into scientific and medical research has promoted the discovery of potential drug targets [3] and facilitated the development of gene therapy approaches for correcting genetic mutations [4].

The discovery of a revolutionary genome editing technology CRISPR/Cas9 in 2014 [5] propelled humanity towards a new era of disease treatment and genetic manipulation. With the ability to precisely modify DNA sequences, CRISPR/Cas9 offers unprecedented opportunities for addressing inherited disorders and potentially reshaping the genetic landscape of future generations [6]. However, the prospect of using genome sequencing and editing technologies to engineer the "ideal" human raises profound ethical, legal, and social implications [7,8], sparking debates about the limits of scientific intervention in human biology and the concept of genetic enhancement.

In this review, we analyzed the human genome of a young woman. Utilizing the dataset obtained, we identified the individual's genetic origin through haplogroup analysis. Moreover, the investigation extended to SNPs associated with phenotypic traits, notably eye and hair color. Furthermore, the study scrutinized the dataset for potential clinical implications, aiming to identify SNPs associated with significant health features. This comprehensive analysis sheds light on the complex interplay between genetic variations and observable traits, offering insights into both ancestral lineage and potential health predispositions.

2. Materials and Methods

In this study, data obtained from the commercial company "Genotek" for Maria Uzun was utilized. Briefly, DNA was extracted from saliva, followed by its genotyping with Illumina HiScan resulting in a VCF file containing single nucleotide polymorphisms (SNPs) for the set of genes of interest.

For the analysis of ancestry and haplogroups, the Haplogrep3 service was employed [9]. Comparison of the obtained data with the human reference genome (GRCh37/hg19) and its visualization were conducted using IGV (v.2.16.1). SNP annotation was performed using Variant Effect Predictor (VEP, v.111) [10] with GRCh38.p14. SNP descriptions were sought on the SNPedia website (https://www.snpedia.com/).

3. Results

Based on the dataset obtained, the origin of the individual was studied based on haplogroups. It was revealed that the genome under study has the H2a2a1 haplogroup, similar to the Cambridge Reference Sequence (CRS) [11].

The H2a2a1 haplogroup, a subclade of haplogroup H, is prevalent among European populations. Haplogroup H itself

is believed to have originated in the Near East around 20,000-25,000 years ago, subsequently migrating into Europe during the Paleolithic period, approximately 15,000-20,000 years ago, following the retreat of the ice sheets [12].

Analysis of single nucleotide polymorphisms (SNPs) related to eye color, according to Hart et al. [13], identified variants associated with the development of non-blue eye color (rs12913832 A/G). However, further investigation was hindered as no other similar variants were identified. Additionally, the analysis suggested a likelihood of dark hair and light skin with the presence of freckles.

Furthermore, the obtained data were analyzed for the presence of certain important clinical features. The results of the analysis are presented in Supplementary Table 1.

4. Discussion

This study examined several clinical variants. Initially, the risk associated with the genetic variant rs6548238 in the *TMEM18* gene, linked to obesity, was investigated, leading to a 0.26 [0.19-0.34] kg/m2 increase in weight and body mass index in C/C homozygotes. However, no such effect was observed in heterozygotes [14]. Therefore, transitioning from homozygosity to heterozygosity could help mitigate risks.

The second investigated risk is associated with hyperbilirubinemia and the SNP rs887829 in the *UGT1A1* gene (NM_000463.3(UGT1A1):c.1024C>T (p.Pro342Ser)). This variant results in decreased enzyme activity (30% of the wild-type), which is necessary for converting unconjugated bilirubin into a water-soluble form and eliminating it from the body. Reduced enzyme function leads to the accumulation of toxic unconjugated bilirubin, which, when circulated in the bloodstream, interacts with cell membrane lipids in various organs and tissues [15]. Substituting this variant with the heterozygous T/C allele could help mitigate disease risks.

The third variant investigated is connected with Transferrin Receptor 2 (*TFR2*) and the rs7385804 SNP. This gene encodes a single-pass type II membrane protein, which belongs to the transferrin receptor-like family. The protein functions in the cellular uptake of transferrin-bound iron and may play a role in iron metabolism, hepatocyte function, and erythrocyte differentiation. The C/A variant is associated with decreased iron status [16]. Conversely, the C/C variant is associated with normal iron absorption and concentration in the blood.

The next variant under consideration is rs1801133 (G/A), which results in the presence of one copy of the C677T allele of *MTHFR*, leading to a 65% efficiency in processing folic acid [17]. The corrected variant, G/G, will result in normal homocysteine levels.

The last variant considered is rs7850258, which is located on chromosome 9 and is 67 kb away from the nearest gene, *FOXE1*, also known as thyroid transcription factor 2. *FOXE1* is an intronless gene crucial for thyroid morphogenesis. This SNP is associated with primary hypothyroidism [18]. Primary hypothyroidism is the most common thyroid disorder, affecting 1-5% of the population, and is characterized by deficiencies of thyroid hormones T3 (triiodothyronine) and T4 (thyroxine) [19]. Changing this variant to A/A would result in slightly lower odds of developing primary hypothyroidism.

References

[1] McVean, G., Spencer, C. C. A. & Chaix, R. Perspectives on human genetic variation from the HapMap project. *PLoS Genet.* **1**, 0413–0418 (2005).

[2] van der Sijde, M. R., Ng, A. & Fu, J. Systems genetics: From GWAS to disease pathways. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842**, 1903–1909 (2014).

[3] Cao, C. & Moult, J. GWAS and drug targets. *BMC Genomics* **15**, 1–14 (2014).

[4] Li, H. *et al.* Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct. Target. Ther.* **5**, (2020).

[5] Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.)* **346**, (2014).

[6] Liu, Y. *et al.* Global chromosome rearrangement induced by CRISPR-Cas9 reshapes the genome and transcriptome of human cells. *Nucleic Acids Res.* **50**, 3456–3474 (2022).

[7] Macer, D. Ethical, legal and social issues of genetically modifying insect vectors for public health. *Insect Biochem. Mol. Biol.* **35**, 649–660 (2005).

[8] Mulvihill, J. J. *et al.* Ethical issues of CRISPR technology and gene editing through the lens of solidarity. *Br. Med. Bull.* **122**, 17–29 (2017).

[9] Schönherr, S., Weissensteiner, H., Kronenberg, F. & Forer, L. Haplogrep 3-an interactive haplogroup classification and analysis platform. *Nucleic Acids Res.* **51**, W263–W268 (2023).

[10] McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 1–14 (2016).

[11] Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA - Nature Genetics. *Nat. Genet.* **23**, 147 (1999).

[12] Roostalu, U. *et al.* Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: The Near Eastern and Caucasian perspective. *Mol. Biol. Evol.* **24**, 436–448 (2007).

[13] Hart, K. L. *et al.* Improved eye- and skin-color prediction based on 8 SNPs. *Croat. Med. J.* **54**, 248–256 (2013).

[14] Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).

[15] Chen, G. *et al.* A UGT1A1 variant is associated with serum total bilirubin levels, which are causal for hypertension in African-ancestry individuals. *npj Genomic Med.* **6**, 1–6 (2021).

[16] An, P. *et al.* TMPRSS6, but not TF, TFR2 or BMP2 variants are associated with increased risk of iron-deficiency anemia. *Hum. Mol. Genet.* **21**, 2124–2131

[17] Clarke, R., Halsey, J., Bennett, D., & Lewington, S. (2011). Homocysteine and vascular disease: Review of published results of the homocysteine-lowering trials. *Journal of Inherited Metabolic Disease*, 34, 83–91.

[18] Denny, J. C., et al. (2011). Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: Using electronic medical records for genome- and phenome-wide studies. *American Journal of Human Genetics*, 89, 529–542.

[19] Eriksson, N., et al. (2012). Novel associations for hypothyroidism include known autoimmune risk loci. *PLoS One*, 7, 1–8.

Supplementary materials

Supplementary Table 1: Analysed SNPs and their characteristics.

snp	chromosome	position	variant	feature	suggested changes
eye color					
rs12913832	15	28365618	A/G	Not blue	-
rs12203592	6	396321	C/T	Primarily in Europeans; likely presence of freckles, brown hair and high sensitivity of skin to sun exposure.	-
rs16891982	5	33951588	G/G	Generally European; Light skin; Possibly an increased risk of melanoma	-
rs12896399	14	92773663	G/G	Brown eyes	-
risks of diseases					
rs6548238	2	634905	C/C	Risk of obesity	C/T
rs887829	2	234668570	T/T	Hyperbilirubinemia I; Gilbert disease	T/C
rs7385804	7	100235970	C/A	Decreased iron status	C/C
rs1801133	1	11856378	G/A	Decreased efficiency in processing folic acid	G/G
rs7850258	9	97786731	G/G	Slightly higher odds of developing primary hypothyroidism.	A/A