BIOINFORMATICS INSTITUTE

Bio-informatics Practice

February 17, 2024

Human upgrade

by Artem Vasilev and Tatiana Lisitsa

Abstract

In recent years, transhumanism has been gaining momentum. This work is aimed at studying SNPs in the genome of a particular person in order to find its clinical predispositions to various diseases, correcting the found variants for favorable ones, and also improving other potentially useful qualities for a human. We were able to define the sex, origin, haplogroup, eye, skin and hair color of a person. Besides, we have found and fixed 3 pathogenic variants and upgraded 2 bening traits in the genome.

Keywords: microarray, CRISPR-Cas9, SNP, genome profiling *Homo sapiens*

Introduction

The rapid advancement of genotyping technologies is opening new horizons for a deeper understanding of genetic characteristics across organisms. This ever-expanding toolkit not only allows scientists to conduct more precise genotyping but also sheds light on the intricate molecular aspects of our genetic heritage. One such technology is DNA microarrays. In tandem with these achievements in genotyping, genome editing technologies are undergoing intensive development. These innovative methods provide researchers with unique opportunities to modify genomes with high precision, offering new perspectives for treating genetic disorders and creating genetically modified organisms with enhanced traits.

A DNA microarray is a solid substrate featuring immobilized probes—oligonucleotides designed to complement specific regions of DNA or RNA. This technology is rooted in the hybridization method, where the DNA sample undergoes amplification, and oligonucleotides tagged with fluorescent markers are introduced during this stage. In the context of RNA analysis, the initial step involves reverse transcription. Subsequently, the sample is applied to the microarray, where the hybridization of DNA molecules with complementary probes takes place. Non-complementary molecules are removed during the array washing process. Following this, the microarray undergoes laser scanning for signal detection. Conclusions about the genotype are derived from regions of the microarray exhibiting fluorescent signals. Oligonucleotides with predefined sequences are immobilized at each site. This technology not only enables the evaluation of individual SNPs but also extends to the assessment of gene expression, detection of structural rearrangements, identification of copy number variations (CNV), and exploration of methylation patterns [1, 2].

Nowadays, one of the most powerful and promising genetic engineering technologies is the CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats-associated 9) system. It was discovered in bacteria in 2012 and subsequently began to be actively studied and used for gene editing [3]. CRISPR-Cas9 is a system that allows bacteria to "remember" and resist viruses. In the process of evolution, bacteria have developed the ability to create short DNA sequences (CRISPR), which contain information about previously encountered viruses. These sequences are then used by Cas9 proteins to recognize and cut viral RNAs, preventing them from entering the cell. The discovery of CRISPR-Cas9 was an important step in the development of gene therapy, since Cas9 is able to resolve double-strand breaks in genes, which allows changes to be made to them. This makes CRISPR-Cas9 a system that can be used for genetic modification or "editing" in living organisms. Using CRISPR-Cas9, you can edit the genome of various organisms, including animals, plants and even humans.

Materials and methods

In this project, we analyzed datasets provided by our teacher - genotyping data performed at 23andMe and at Genotek. 23andMe analyzed 272866 SNPs and Genotek analyzed 147106 SNPs. There were 53188 SNPs common in the two datasets, but genotype matched for 53181 SNPs, hence for the common 7 SNPs the genotyping results differed. Positions in the vcf file are

specified for GRCh37 human genome assembly. We analyzed SNPs lying on the Y chromosome and in the mitochondrial DNA using the MorleyDNA and mtDNA Haplogroup Analysis web services, respectively. Then we analyzed SNPs associated with eye, skin, and hair color using material from the article "Improved eye-and skin-color prediction based on 8 SNPs" [4] and the HIris-Plex-S. To annotate the nucleotide sequence variants, we used the web version of VEP [5] and desktop version OpenCRAVAT. Clinical significance was assessed using ACMG criteria [6]. We used the Franklin by genoox service to clarify information about the nucleotide sequence variant.

Results

The most probable haplogroup identified by Y-chromosome is R1a1a. Haplogroup R1a1a, also referred to as R-M17, stands out as a prominent subgroup within the broader haplogroup R1a. Widely dispersed, haplogroup R1a is prevalent across Europe, Central and East Asia, as well as certain regions in northern India. Specifically, Haplogroup R1a1a is frequently linked with Indo-European populations and languages; however, its presence extends to diverse ethnic groups worldwide [7].

Best mtDNA Haplogroup Match was H(T152C). Haplogroup H is one of the most common haplogroups among European populations. It is also found in some peoples of North Africa, the Middle East, and Western Asia. Haplogroup H is quite ancient, representing one of the oldest mtDNA lineages in Europe. Its origin is typically traced back to the Upper Paleolithic era [8].

Using the method described in this article and data used, it was not possible to accurately determine eye color. However, we can say with confidence that it is not blue, not brown and not green.

However, using the HIris-Plex-S web-tool allowed us to more accurately determine color characteristics (Table 1).

Table 1: Color characteristic

Characteristic	**Variants**	**p-value**
Eye	Blue	0.033239463
	Intermediate	0.10772317
	Brown	**0.859037367**
Hair	Blond	0.019648391
	Brown	**0.582082833**
	Red	0.000735681
	Black	0.397533095
Skin	VeryPale	0.020355307
	Pale	0.171512807
	Intermediate	**0.754441269**
	Dark	0.049819315
	DarktoBlack	0.003871302

Most likely, this man has an intermediate skin color with dark brown hair and brown eyes (based on the results of the first analysis, it makes sense to assume that the eye color is also intermediate).

The man is least likely to be black with red hair and blue eyes.

Based on ACMG criteria and in the absence of personal and family history information, we did not identify pathogenic or likely pathogenic clinically significant variants that could explain the phenotype. Nevertheless, we selected a few variants that might have manifested phenotypically in this individual or their offspring, either for potential correction or simply for desirable traits.

Table 2: Relevant SNPs

Description	SNP ids	Current GT	Suggested GT
Pathogenic fixes			
Adrenal hyperplasia, AR, Hyperandrogenism, AR	i5005436	C/T	C/C
Inflammatory bowel disease, susceptibility to Systemic lupus erythematosus	rs2004640	G/T	G/G
Modifier of Coronary artery disease susceptibility to Mycobacterium tuberculosis susceptibility to Spina bifida	rs1024611	A/G	A/A
Upgrades			
Body resistance to direct vacuum exposure improved	rs10489156	A/G	C/T
	rs7541616	./.	G/G
No problems with understanding recursion	rs17162330	C/T	G/G
	rs17162339	C/C	G/G
	rs3010109	G/G	T/T

Discussion

Variant i5005436 is located in the *CYP21A2* gene, bearing the HGVS name c.1069C>T, this is a missense mutation. The substitution of cytosine by thymine in this region leads to a protein alteration, where arginine is replaced by tryptophan at codon 357 (p.Arg357Trp). The CYP21A2 gene encodes steroid 21-hydroxylase, playing a crucial role in the metabolism of steroid hormones in the adrenal cortex. Homozygous or compound heterozygous mutations in this gene result in 21-hydroxylase deficiency, leading to the manifestation of either Adrenal Hyperplasia or non-classical Hyperandrogenism phenotypes.

The rs2004640 variant affects the interferon regulatory factor 5 gene (*IRF5*) and is designated as HGVS c.-12+2G>T. Interferon regulatory factors (IRFs) are transcription factors that mediate inflammatory signaling pathways. Damage to this gene can lead to susceptibility to inflammatory and autoimmune diseases.

The rs10246110 variant is located in the *CCL2* regulatory region and is named HGVS c.-2581A>G. The CCL2 protein (monocyte chemotactic protein-1, a member of the small inducible gene (SIG) family, plays a role in the recruitment of monocytes to sites of injury and infection.

For more detailed information about genes and SNPs associated with various diseases and phenotypic characteristics, it is recommended that you consult a geneticist or human geneticist.

The proposed improvements (vacuum resistance and recursion understanding) were taken from this repository (GitHub). The authors do not guarantee their performance.

P.S. I would like to clarify my stance – I strongly advise against taking these tests for entertainment or educational purposes. If you are not an expert in genotyping and result interpretation, or in the field of medical genetics, it will be challenging to assess genotyping results accurately. Unexpected outcomes may arise, leading to additional stress. Furthermore, I believe that such tests can potentially harm the reputation of the medical genetics specialty.

Supplementary materials

GitHub

References

[1] Marco Wiltgen and Gernot P Tilz. "DNA microarray analysis: principles and clinical impact". In: *Hematology* 12.4 (2007), pp. 271–287.

[2] Yiping Shen and Bai-Lin Wu. "Microarray-based genomic DNA profiling technologies in clinical molecular diagnostics". In: *Clinical chemistry* 55.4 (2009), pp. 659–669.

[3] Martin Jinek et al. "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity". In: *Science* 337.6096 (2012), pp. 816–821. DOI: 10.1126/science.1225829. eprint: https://www.science.org/doi/pdf/10.1126/science.1225829. URL: https://www.science.org/doi/abs/10.1126/science.1225829.

[4] Katie L Hart et al. "Improved eye-and skin-color prediction based on 8 SNPs". In: *Croatian medical journal* 54.3 (2013), pp. 248–256.

[5] William McLaren et al. "The ensembl variant effect predictor". In: *Genome biology* 17.1 (2016), pp. 1–14.

[6] Sue Richards et al. "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology". In: *Genetics in medicine* 17.5 (2015), pp. 405–423.

[7] Peter A Underhill et al. "The phylogenetic and geographic structure of Y-chromosome haplogroup R1a". In: *European Journal of Human Genetics* 23.1 (2015), pp. 124–131.

[8] Eva-Liis Loogväli et al. "Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia". In: *Molecular Biology and Evolution* 21.11 (2004), pp. 2012–2021.