What information can my genome provide?

Michael Greenberbg, Polina Vaganova
Bioinformatics Institute, Saint Petersburg, Russia

Abstract.
Small differences in our genome can reveal a lot about various traits of an individual, ranging from their ancestry and physical appearance to predispositions to certain diseases. Such information holds the potential to shed light on three big questions that have intrigued humanity for ages: Where do we come from? Who are we? And where are we going? In our study, we aimed to address all three of these questions using real person genome data.

Keywords: human genome, SNPs, CRISPR-Cas9, transhumanism .

1 Introduction

Where do we come from? Who are we? Where are we going?

Identification of single nucleotide polymorphisms (SNPs) and other mutations in the human genome provides valuable information about various traits of an individual, which can be utilized for both cultural and medical purposes. Despite sequencing technologies become increasingly affordable, methods such as DNA microchips, which operate based on DNA-DNA hybridization, enable cheaper, faster, and simultaneous identification of thousands of polymorphisms in human genomes.[1,2]

At the same time, technologies that are capable to change our answer to the question "Where are we going?" are rapidly advancing. For instance, the CRISPR-Cas9 system, derived from prokaryotic immune mechanisms, is a powerful genome editing tool that allows researchers to precisely modify DNA sequences,[3] offering potential solutions for addressing undesirable genetic traits.

The objective of our study is to determine the ancestry and physical traits of a real person based on genetic polymorphisms, as well as to identify SNPs associated with the risk of health issues and parts of the genome that can be improved - potential targets for treatment using CRISPR-Cas9.

2 Materials and methods

In this study we analyzed Genotek raw data for one of the authors. Raw data was downloaded from the Genotek portal in the both vcf and 23andme formats. It contains information about subject's SNPs, including its rsID, chromosome position and resulting genotype.

Identification of SNPs in the maternal (mtDNA) that distinguish haplogroup was conducted using James Lick's mtHap utility.

We used VEP (Variant Effect Predictor) web version to annotate SNPs in the given vcf file. We extracted information about SNPs related to risk factors from the VEP output. Additional information about these SNPs was obtained from dbSNP database.

3 Results

In humans mitochondrial DNA is maternally inherited. So in out research we analyzed SNPs in mitochondrial DNA to establish author's maternal ethnicity. In the Table 1 we summarised match results and revealed maternal haplogroups.

Table 1: mtDNA haplogroup with high matching level and information about match parameters.

Haplogroup	Matches	Mismatches	Extras
H2a2a	1	0	6
H2a2a1	0	0	7
H2a	2	2	5
H2a2a1d	0	0	7
H2a2a1b	0	0	7

According to our findings most possible subject's mtDNA haplogroup is H2a2a1. Estimated phylogenetic path of human migrations with this haplogroup is shown in the Figure 1.

Fig 1: Schematic illustration of human migrations with H2a2a1 halogroup, according to Genomed.

We analyzed SNPs markers responsible for the eye color.[4] Obtained data collected in the Table 2.

Table 2: SNPs that determine eye color, subject's genotype and probable eye color

SNP	Genotype	Probable eye color
rs12913832	G/G	Not brown
rs12203592	C/T	Green or blue
rs16891982	G/G	Not green

We revealed all clinically relevant SNPs, which associated with risk factors. In total, we

found four SNPs, connected with predispositions to different diseases. These SNPs summarized in the Table 3.

Table 3: revealed SNPs and probable disease, associated with them.

SNP	Susceptibility to to disease
rs6548238	Obesity
rs104894130	Oculocutaneous albinism type 2
rs77375493	Budd-Chiari syndrome
rs28357675	Leigh syndrome

One of our goals was to propose some changes of genome which could make life better. We inspected dbSNP database and PubMed to find data about such possible improvements. In the Table 4 you can see useful features, connected to genes manipulations, which were selected by us as the most promising ones.

Table 4: Probable health improvements and gene manipulations we need to get them.

Potential trait	Gene	Manipulation to do
reduced levels of total and LDL cholesterol	APOE	rs429358 (C $\rightarrow$ T)
increase ability to fighting oxidative stress	FOXO3A	rs2802292 (G $\rightarrow$ T)
improved memory performance	KIBRA	rs17070145 (C $\rightarrow$ T)
improved memory performance	CLSTN2	rs6439886 (C $\rightarrow$ T)

4 Discussion

In our study, we revealed information about subject's ethnicity, eye color, and potential health risks. Finally, we suggested some changes, which may be useful for anyone (such as improved mental abilities and overall health).

Through haplogroup analysis, we identified a connection with the H2a2a1 haplogroup, predominantly found among European populations (according to FamilyTreeDNA). We didn't found found any paternal haplogroups, thus subject's biological sex was determined correctly as a female.

Furthermore, eye color prediction, based on SNPs analysis accurately indicated a light coloration, which is consistent with the the subject's real phenotype.

SNPs annotation analysis show four subject's potential health risks. Only two out of four identified health risks were considered life-threatening: Leigh and Budd-Chiari syndromes. The other two are related to obesity risk, which can be quite controlled by subject's lifestyle and oculocutaneous albinism that may be expressed only in the subject's descendants. Interestingly, that previous genetic analysis, performed by other specialists, revealed list of predispositions to diseases, which is significantly different from the one, presented here.

Additionally, we proposed some enhancements that could be introduced into the subject's genome. First of all, these include mutations in the APOE gene (rs429358) and FOXO3A

gene (rs2802292). Researches had demonstrated that these mutations are associated with significantly increased chances of achieving extreme longevity,[5] reduced risk of death,[6] healthy cognitive aging, and offer protective effects against several age-related diseases.[7]

Another improvements involve introducing specific mutations in the genes encoding the KIBRA and CLSTN2 proteins (rs17070145 and rs6439886, respectively). These proteins are expressed in the brain, and particular mutations in them have been linked to enhanced memory function.[8]

All of this enhancements can be introduced using CRISPR-Cas9 system. However, nowadays it is not used in medical practice. Still with recent technologies we can continuing to reveal hypothetical changes for the future and answers to three main questions:

Where do we come from? Who are we? Where are we going?

References

1 P. Muir, S. Li, S. Lou, *et al.*, "The real cost of sequencing: scaling computation to keep pace with data generation," *Genome Biology* **17** (2016).

2 T. LaFramboise, "Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances," *Nucleic Acids Research* **37**, 4181–4193 (2009).

3 A. Khurana, N. Sayed, V. Singh, *et al.*, "A comprehensive overview of crispr/cas 9 technology and application thereof in drug discovery," *Journal of Cellular Biochemistry* **123**, 1674–1698 (2022).

4 K. L. Hart, S. L. Kimura, V. Mushailov, *et al.*, "Improved eye- and skin-color prediction based on 8 snps," *Croatian Medical Journal* **54**, 248–256 (2013).

5 A. Gurinovich, S. L. Andersen, A. Puca, *et al.*, "Varying effects of apoe alleles on extreme longevity in european ethnicities," *The Journals of Gerontology: Series A* **74**, S45–S51 (2019).

6 P. Sebastiani, A. Gurinovich, M. Nygaard, *et al.*, "Apoealleles and extreme human longevity," *The Journals of Gerontology: Series A* **74**, 44–51 (2018).

7 M. Revelas, A. Thalamuthu, C. Oldmeadow, *et al.*, "Review and meta-analysis of genetic polymorphisms associated with exceptional human longevity," *Mechanisms of Ageing and Development* **175**, 24–34 (2018).

8 A. Papassotiropoulos, D. A. Stephan, M. J. Huentelman, *et al.*, "Common kibra alleles are associated with human memory performance," *Science* **314**, 475–478 (2006).