Genome analysis

H+, or how to build a perfect human.

Bogdanova Liliia[1], Vlasenko Daniil[2], Zhurakovskaya Anya[3]

[1]Sechenov University, Specialist in Bioengineering and Bioinformatics, Moscow, Russia, E-mail: liliia.bgdnv@gmail.com
[2]St. Petersburg State University, School of Mathematics and Mechanics, St. Petersburg, Russia, E-mail: vlasnko.daniil.vl@gmail.com
[3]Moscow State University

Abstract

This article discusses the use of SNP (single nucleotide polymorphism) chips and CRISPR-Cas9 systems for genetic analysis and genome editing. SNP chips are DNA microarrays that analyze genetic variation across the genome and are effective for examining common genetic variation. However, they struggle to genotype uncommon genetic variations. On the other hand, CRISPR-Cas9 systems are effective for genome editing as they alter genomic DNA at specific target sites. The article suggests ten SNP changes to improve a person's quality of life and prevent potential diseases, including changes related to cardiovascular diseases and type 2 diabetes. Additionally, the article proposes five changes that will generally make the person's life easier.

Supplementary information: https://github.com/Daniil-Vlasenko/IBBioinformaticsWorkshop/blob/main/Project%205

1 Introduction

SNP chips are DNA microarrays used to analyze genetic variation at the many millions of unique places across the genome. The investigation of human disease's genetic composition has been significantly impacted by the introduction of microarray technology. They were first created to assess single nucleotide polymorphisms (SNPs) that are prevalent in the community and affect more than one in one hundred persons. SNP chips have demonstrated to be highly effective for examining common genetic variation, which may be used to determine ancestry and propensity to a variety of complicated multifactorial disorders, including type 2 diabetes. The genetics community is aware that SNP chips struggle to genotype uncommon genetic variations because they rely heavily on data grouping. [1]

The availability of the ability to sequence the genome leads to the idea of genome editing, for which CRISPR-Cas9 systems are currently used. Genome editing, also known as insertion, deletion, and replacement of DNA, is the alteration of genomic DNA at a specific target site in a wide range of cell types and organisms. This results in the inactivation of target genes, the acquisition of novel genetic traits, and the correction of pathogenic gene mutations. CRISPR-Cas systems have emerged as the most popular genome editing tool in molecular biology labs all over the world because of its benefits of simple design, low cost, high efficiency, strong reproducibility, and quick cycle. Zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the RNA-guided CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) nucleases systems are the three most common genome editing methods available today. Most bacteria and archaea have CRISPR-Cas, an adaptive immune system that protects them against phages, viruses, and other foreign genetic material. It is made up of a group of CRISPR-associated genes that encode Cas proteins with endonuclease activity and CRISPR repeat-spacer arrays, which may be further translated into CRISPR RNA and trans-activating CRISPR RNA (tracrRNA). When foreign genetic elements penetrate prokaryotes, Cas proteins can break the invaders' DNA into brief bits, which are subsequently incorporated into the CRISPR array as new spacers. When the same invader attacks again, crRNA will detect it right away and couple with the foreign DNA, which directs Cas protein to break specific foreign DNA target regions, defending the host. [2]

2 Methods

We used the 23andMe results of our teacher [3] and used plink [4] to remove all SNPs corresponding to deletions and insertions to make the file compatible with annotation tools: **plink –23file SNP_raw_v4_Full_20170514175358.txt –recode vcf –out snps_clean –output-chr MT –snps-only just-acgt**.

Then we established the probable ethnicity of our teacher by identifying maternal (mtDNA) with James Lick's mtHap utility [5] and paternal (Y chromosome) haplogroups with MorleyDNA.com Y-SNP Subclade Predictor [6].

To determine eye and skin color we analyzed 8 SNP's significantly associated with these features [7] using Integrative Genomics Viewer [8]. Finally, we used Variant Effect Predictor [9] with dbSNP database [10] to annotate all SNPs and select clinically relevant ones.

3 Results

James Lick's mtHap utility determined that the maternal haplogroup can be H(T152C) (most likely), H1(T152C), H, H16(T152C), H69, H46, H52, H3(T152C), H9. MorleyDNA.com Y-SNP Subclade Predictor determined that the paternal haplogroup can be R1a1a (most likely), M3, R1b, N1a.

Table 1 contains all SNPs responsible for eye color. Table 2 contains all SNPs that are related to increased risk for heart disease and diabetes.

Table 1. SNPs responsible for eye color.

rsid	SNP
rs12913832	AG
rs16891982	GG
rs6119471	—
rs12203592	CT
rs16891982	CG
rs12896399	GG

Table 2. SNPs that are related to increased risk for heart disease.

rsid	SNP	description	replacement
rs10757274	AG	1.3x increased risk for heart disease	AA
rs4977574	AG	1.8x increased risk for high blood pressure	AA
rs4961	GT	1.8x increased risk for high blood pressure	GG
rs699	AG	increased risk of hypertension	TT
rs13266634	CT	increased risk for type-2 diabetes	TT
rs1799987	-	HIV resistant mutation	
rs1805007	CC	increased response to anesthetics	TT
rs4988235	GG	can digest milk	TT
rs182549	CC	can digest milk	CT
rs602662	AG	Higher vitamin B12 levels	AA

4 Discussion

To improve this person's quality of life and prevent some potential diseases we suggest 10 SNP changes. While analyzing the results of clinically relevant SNPs we noticed, that this particular person has increased risks of cardiovascular diseases and type 2 diabetes. These diseases are also quite common in the population... <>

The rs10757274 SNP is described as SNP strongly linked to coronary heart disease, noncardioembolic stroke, and early myocardial infarction (heart attack) [11, 12, 13]. Although we do not know the exact mechanism of action of affected protein, the evidence that this SNP plays an important role in such diseases is abundant. Similarly, the rs4977574 SNP is also shown to be connected to an increased susceptibility to coronary artery disease in several Genome-wide association (GWAS) studies [14, 15, 16]

SNPs rs4961 and rs699 are linked to an increased risk for high blood pressure (hypertension) which, if goes unchecked, is a significant risk factor for stroke, coronary heart disease, heart failure, etc. The rs699 SNP is a variation in ADD1 gene, which is a gene that encodes adducin protein. Adductin is a cytoskeleton protein that promotes binding spectrin-actin binding and controls actin polymerization rates. It is shown that mutated adducin increases Na-K pump activity and impairs Na-K pump endocytosis which leads to abnormal cell sodium handling and hypertension [17]. The rs699 SNP encodes a functional change in the AGT gene which encodes angiotensinogen protein, the precursor to angiotensin. Angiotensin is a hormone that regulates blood pressure by narrowing blood vessels. The variant GT of rs699 is connected to an increased level of angiotensin in the blood which eventually causes hypertension [18].

We also suggest making one change of rs13266634 SNP in SLC30A8 gene, which encodes zinc transporter protein and is expressed only in insulin-producing cells. As shown in GWAS studies, this SNP is significantly associated with type-2 diabetes [19, 20].

Additionally, we propose 5 changes, that will make the life of the person generally easier. First, we suggest adding a mutation to rs1799987 in the CCR5 gene, which makes a person HIV-resistant. Secondly, we can change rs1805007 SNP and not only make our person's hair red but also make them more responsive to anesthetics, which is an overall useful feature. We also would like for our subject to be able to enjoy a bigger variety of food. Therefore we suggest changing rs4988235 and rs182549, so the person would be able to digest milk easily. Finally, we recommend changing rs602662 SNP, which is associated with higher vitamin B12 levels, not only because it can be quite beneficial for someone with a risk of cardiovascular disease and stroke [21].

References

[1]Weedon M.N., Jackson L., et al. "Use of SNP chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation." BMJ 2021. doi:10.1136/bmj.n214.

[2]Xu Y., Li Z. "CRISPR-Cas systems: Overview, innovations and applications in human disease research and gene therapy." Computational and structural biotechnology journal 18 (2020). https://doi.org/10.1016/j.csbj.2020.08.031.

[3]23andMe data; Available from: https://drive.google.com/open?id=1QJkwJe5Xl_jSVpqdTSNXP7sqlYfI666j.

[4]Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, de Bakker PIW, Daly MJ & Sham PC (in press); PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics.

[5]James Lick's mtHap utility A tool for checking and assigning mtDNA haplogroups on an individual basis. Available from: https://dna.jameslick.com/mthap.

[6]MorleyDNA.com Y-SNP Subclade Predictor: extract Y-DNA from an autosomal test. Available from: https://ytree.morleydna.com/extractFromAutosomal.

[7]Hart, Katie L. et al. "Improved eye- and skin-color prediction based on 8 SNPs." Croatian Medical Journal 54 (2013): 248 - 256.

[8]ames T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 (2011).

[9]McLaren, W., Gil, L., Hunt, S.E. et al. The Ensembl Variant Effect Predictor. Genome Biol 17, 122 (2016). https://doi.org/10.1186/s13059-016-0974-4

[10]Sherry,S.T., Ward,M. and Sirotkin,K. (1999) dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. Genome Res., 9, 677–679.

[11]Luke, May M et al. "Polymorphisms associated with both noncardioembolic stroke and coronary heart disease: vienna stroke registry." Cerebrovascular diseases (Basel, Switzerland) vol. 28,5 (2009): 499-504. doi:10.1159/000236914

[12]Paynter, Nina P et al. "Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3." Annals of internal medicine vol. 150,2 (2009): 65-72. doi:10.7326/0003-4819-150-2-200901200-00003

[13]Nina P. Paynter, Daniel I. Chasman, Julie E. Buring, et al. Cardiovascular Disease Risk Prediction With and Without Knowledge of Genetic Variation at Chromosome 9p21.3. Ann Intern Med.2009;150:65-72. [Epub 20 January 2009]. doi:10.7326/0003-4819-150-2-200901200-00003

[14]Denny, Joshua C et al. "Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data." Nature biotechnology vol. 31,12 (2013): 1102-10. doi:10.1038/nbt.2749

[15]Beckie, Theresa M et al. "The relationship between polymorphisms on chromosome 9p21 and age of onset of coronary heart disease in black and white women." Genetic testing and molecular biomarkers vol. 15,6 (2011): 435-42. doi:10.1089/gtmb.2010.0222

[16]Saade, Stephanie et al. “Large scale association analysis identifies three susceptibility loci for coronary artery disease.” PloS one vol. 6,12 (2011): e29427. doi:10.1371/journal.pone.0029427

[17]Bianchi, Giuseppe et al. “Adducin polymorphism: detection and impact on hypertension and related disorders.” Hypertension (Dallas, Tex. : 1979) vol. 45,3 (2005): 331-40. doi:10.1161/01.HYP.0000156497.39375.37

[18]Nakajima, Toshiaki et al. “Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations.” American journal of human genetics vol. 70,1 (2002): 108-23. doi:10.1086/338454

[19]Staiger, Harald et al. “Polymorphisms within novel risk loci for type 2 diabetes determine beta-cell function.” PloS one vol. 2,9 e832. 5 Sep. 2007, doi:10.1371/journal.pone.0000832

[20]Omori, Shintaro et al. “Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population.” Diabetes vol. 57,3 (2008): 791-5. doi:10.2337/db07-0979

[21]Hung J, Beilby J P, Knuiman M W, Divitini M. Folate and vitamin B-12 and risk of fatal cardiovascular disease: cohort study from Busselton, Western Australia BMJ 2003; 326 :131 doi:10.1136/bmj.326.7381.131