H+, or how to build a perfect human

Abstract

SNP arrays are widely used to detect polymorphism within a population. This data can be used to trace human migrations and study evolutionary relationships, as well as to determine individual phenotype features. SNP-based genetic linkage analysis can be applied to mapping disease loci and find disease susceptibility genes in individuals. In the future CRISPR-Cas9 and other gene-editing technologies might be able to target these SNPs to correct genetic defects at the source, offering potential treatment for genetic disorders. The goal of this practical work was to examine unprocessed 23andMe data of unknown persons, annotate SNPs, reveal phenotypic traits embedded within the raw data, and apply modifications to specific genetic positions to achieve preferred qualities.

Introduction

Single Nucleotide Polymorphisms (SNPs) are variations at a single position in DNA among individuals. Detecting these variations might be helpfull for understanding genetic diversity, disease predisposition and potential drug response [1].

Microarrays, also known as DNA chips, are a popular technology for SNP detection, offering cost-effective methods for analyzing thousands to millions of SNPs across many samples simultaneously [2,3]. The principle of operation involves the hybridization of labeled with fluorescent dyes DNA fragments from a sample to complementary DNA probes fixed on a solid surface. The presence or absence of specific SNPs is determined by the pattern of fluorescence on the chip, allowing for high-throughput analysis of genetic information. The appearance of DNA microarrays enabled genome-wide association studies (GWAS), which have identified numerous genetic variants associated with complex diseases and traits [4].

CRISPR-Cas9 is a technology which was developed on the bases of bacteria natural immune defense mechanisms and which may be applied to target genome editing of different organisms, including humans. By designing guide RNAs (gRNAs) that match the target sequence, the Cas9 enzyme is directed to the desired location, where it creates a double-strand break in the DNA. This break can then be repaired through the cell's natural mechanisms, allowing for the insertion, deletion, or replacement of genetic material [5].

The implication of CRISPR-Cas9 technology may be profound: from treating genetic disorders, to combating infectious diseases by editing the genomes of pathogens. However, there are still unresolved problems such as immunogenicity, effective delivery systems, off-target effects, and ethical issues which are a barrier to human clinical application [6].

The aim of this practice was to analyze the raw 23andme data and to establish haplogroups, annotate SNPs, identify phenotypic traits from the raw data and make changes in some positions to gain desired characteristics.

Methods

The genetic information for Manu Sporny from the Illumina Omnieexpress Plus Genotyping Beadchip was studied in the analysis [7]. Raw data was provided in 23andMe format. 23andMe file was converted to vcf format by PLINK v1.90b7.2 [8]. All SNPs corresponding to deletions and insertions were removed (--snps-only just-acgt).

Human mitochondrial DNA haplogroup was defined through James Lick's mtHap utility version 17.0 [9] and Y-haplogroup was defined through YSEQ Clade Finder [10]. The SNPs annotation was based on the databases dbSNP [11], SNPedia [12], ClinVar [13]. Clinical relevant ones were obtained with Variant Effect Predictor [14] and SnpSift [15].

Results

The analyzed data allowed us to determine some characteristics of this person, such as eye color, possible origin, and the presence of any diseases. Manu's eye color is not blue according to the study [16]. The result was clarified by the SNP rs12913832, which is associated with brown eyes. [17].

M6a1 was defined as mitochondrial DNA haplogroup. This haplogroup presumably originated in Southeast Asia around 30 000 - 35 000 years ago. It is a descendant of haplogroup M6, which, in turn, descended from haplogroup M [18]. This line is found the most among the population of India and also among the United Arab Emirates, Indonesia, Iran, Saudi Arabia, etc [19, 20].

J-FGC75679 was defined as Y-haplogroup. This haplogroup is found in representatives of Tunisia, Latvia, Lithuania, Poland and some other countries [21]. The J-FGC75679 line split from the ancestor J-BY32777 around 1250 BCE. The last common ancestor of the line dates back to 950CE in the Middle Ages [22]. Thus, it can be assumed that this subject may have a mother from India, and a father from Tunisia.

The annotation of a vcf file containing SNP data with information from the ClinVar database using SnpSift led to the following results: 16 found variants were of unknown significance, 2 were benign and 10 - likely benign(suppl. table 1.). And there were no pathogenic variants.

Comparison with a GWAS catalog after filtering only those factors from a GWAS dataset that suggest a strong positive association, the following associations were detected:

rs11708202 0/1 Alcohol_consumption_x_playing_computer_games_interaction

rs6825410 0/1 Diastolic_blood_pressure_in_combination_therapy__beta_blocker
rs11739417 0/1 Youthful_appearance_self-reported_
rs9356704 0/1 Ulcerative_colitis
rs6980713 0/1 Glaucoma__primary_open-angle_
rs2497219 0/1 Schizophrenia
rs17782124 0/1 Macroalbuminuria_in_type_1_diabetes

The search across VEP (Variant Effect Predictor) didn't reveal any variants of any clinical significance.

Discussion

Some changes in the genome have been proposed that can be carried out using the CRISPR-cas system for raising the quality of life (Table 1).

Table 1. Proposed changes

rsid	**initial variant**	**fixed variant**	**explanation**	**reference**
rs2802292	G/T	G/G	Increased longevity	https://www.snpedia.com/index.php/Rs2802292
rs6983267	G/T	T/T	Decreasing risk of prostate cancer	https://www.snpedia.com/index.php/Rs6983267
rs4680	G/A	A/A	Improving memory and attention tasks	https://www.snpedia.com/index.php/Rs4680
rs6152	G/G	A/A	Reducing the likelihood of baldness	https://www.snpedia.com/index.php/Rs6152(G;G)
rs1815739	T/T	C/C	Improving muscles performance	https://www.snpedia.com/index.php/Rs1815739/
rs17822931	C/T	T/T	Changing the type of earwax	https://www.snpedia.com/ind

			and body odor	ex.php/Rs17822931

rs6152

Variant rs6152 is located in X chromosome in the androgen receptor (AR) gene. The association of the AR with male pattern baldness has been demonstrated [23]. One possible explanation for this association is the interaction between the male hormone dihydrotestosterone (DHT) and the hair follicles on the scalp. DHT binds to AR, which triggers a cascade of events leading to the miniaturization of hair follicles. Miniaturization of hair follicles leads to thinning and shortening of hair. Eventually, the hair follicles stop producing hair, which leads to baldness. Mutations in the AR gene can lead to increased sensitivity of hair follicles to DHT and lead to baldness [24].

Rs1815739

Rs1815739 is responsible for muscles performance and athletic abilities. This refers to the *ACTN3* gene on the eleventh chromosome. This SNP determines the presence or absence of the muscle protein alpha-actinin-3. It has been shown that the T/T genotype is insufficiently represented in strength athletes [25]. This variant encodes a stop codon, which leads to a deficiency or complete absence of alpha-actinin-3, which apparently impairs muscle function.

rs17822931

SNP rs17822931 is located in the *ABCC11* gene on the sixteenth chromosome. This causes body odor, as well as determines the type of earwax and lipid secretion. The product of the *ABCC11* gene is an ATP-binding cassette transporter sub-family C member 11, which is involved in transport through apical membranes. It has been shown that ABCC11 is expressed and localized in the apocrine sweat glands and plays a key role in the secretion of odorous substances and their precursors, which, when exposed to bacteria, cause odor [26].

rs2802292

The subject is heterozygous (G/T) at position 109015211 on chromosome 6. This SNP happens in FOXO3 gene in the intron. This gene is involved in the regulation of oxidative stress, insulin sensitivity, and cellular apoptosis. SNPs in the FOXO3 gene have been linked to longevity in various populations.

If we were to change his genotype to homozygous GG at this position, the subject would be 1.5 to 2.7 times more likely to live to 100.

rs6983267

The heterozygous (G/T) genotype at position 128482487 on chromosome 8 has an increased risk of prostate cancer; risk genotypes yield an odds ratio for developing prostate cancer of 1.37 (CI: 1.18-1.59, p=3.4-10e-5) and may account for

22.2% of population. This SNP belongs to two genes: CCAT2 on a plus strand and presents a non coding transcript variant, and CASC8 gene on a minus strand where it can be part of either genic downstream transcript variant or intron variant. If we correct this genotype to homozygous T/T, the risk of cancer will become normal.

rs4680

rs4680 (Val158Met) is a well studied SNP in the COMT gene. The COMT gene codes for the COMT enzyme, which breaks down dopamine in the brain's prefrontal cortex. The wild-type allele is a (G), coding for a valine amino acid; the (A) substitution polymorphism changes the amino acid to a methionine. This alters the structure of the resultant enzyme such that its activity is only 25% of the wild type.

The heterozygous variant has intermediate dopamine levels. If we change the genotype to homozygous A/A, our subject will acquire advantage in memory and attention tasks.

Literature

1. Kim S, Misra A. SNP genotyping: technologies and biomedical applications. Annu Rev Biomed Eng. 2007;9:289-320. doi: 10.1146/annurev.bioeng.9.060906.152037.
2. Van Asselt AJ, Ehli EA. Whole-Genome Genotyping Using DNA Microarrays for Population Genetics. Methods Mol Biol. 2022;2418:269-287. doi: 10.1007/978-1-0716-1920-9_16.
3. Lam CW, Lau KC, Tong SF. Microarrays for personalized genomic medicine. Adv Clin Chem. 2010;52:1-18. doi: 10.1016/s0065-2423(10)52001-8.
4. Yang TH, Kon M, DeLisi C. Genome-wide association studies. Methods Mol Biol. 2013;939:233-51. doi: 10.1007/978-1-62703-107-3_15. PMID: 23192550.
5. Ran, F., Hsu, P., Wright, J. et al. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281–2308 (2013). https://doi.org/10.1038/nprot.2013.143
6. Uddin F, Rudin CM, Sen T. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. Front Oncol. 2020 Aug 7;10:1387. doi: 10.3389/fonc.2020.01387. PMID: 32850447; PMCID: PMC7427626.
7. Manu Sporny's genetic information https://github.com/msporny/dna
8. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4 www.cog-genomics.org/plink/1.9/
9. https://dna.jameslick.com/mthap/
10. http://predict.yseq.net/clade-finder
11. 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

12. Cariaso M, Lennon G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. Nucleic Acids Res. 2012 Jan;40(Database issue):D1308-12. doi: 10.1093/nar/gkr798.
13. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014 Jan;42(Database issue):D980-5. doi: 10.1093/nar/gkt1113.
14. McLaren, W., Gil, L., Hunt, S.E. et al. The Ensembl Variant Effect Predictor. Genome Biol 17, 122 (2016). doi: 10.1186/s13059-016-0974-4
15. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet. 2012 Mar 15;3:35. doi: 10.3389/fgene.2012.00035.
16. Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E. Improved eye- and skin-color prediction based on 8 SNPs. Croat Med J. 2013 Jun;54(3):248-56. doi: 10.3325/cmj.2013.54.248.
17. https://www.snpedia.com/index.php/Rs12913832
18. Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, Bandopadhyay SS, Barua P, Barik SS, Basu D, Kiran U, Gangopadhyay P, Sahani R, Prasad BV, Gangopadhyay S, Lakshmi GR, Ravuri RR, Padmaja K, Venugopal PN, Sharma MB, Rao VR. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PLoS One. 2009 Oct 13;4(10):e7447. doi: 10.1371/journal.pone.0007447.
19. Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet. 2004 Aug 31;5:26. doi: 10.1186/1471-2156-5-26.
20. https://www.familytreedna.com/public/mt-dna-haplotree/M
21. https://www.yfull.com/branch-info/J-FGC75679
22. https://discover.familytreedna.com/y-dna/J-FGC75679/story
23. Ellis JA, Stebbing M, Harrap SB. Polymorphism of the androgen receptor gene is associated with male pattern baldness. J Invest Dermatol. 2001 Mar;116(3):452-5. doi: 10.1046/j.1523-1747.2001.01261.x.
24. Lolli F, Pallotti F, Rossi A, Fortuna MC, Caro G, Lenzi A, Sansone A, Lombardo F. Androgenetic alopecia: a review. Endocrine. 2017 Jul;57(1):9-17. doi: 10.1007/s12020-017-1280-y.
25. Roth SM, Walsh S, Liu D, Metter EJ, Ferrucci L, Hurley BF. The ACTN3 R577X nonsense allele is under-represented in elite-level strength athletes. Eur J Hum Genet. 2008 Mar;16(3):391-4. doi: 10.1038/sj.ejhg.5201964.

26. Martin A, Saathoff M, Kuhn F, Max H, Terstegen L, Natsch A. A functional ABCC11 allele is essential in the biochemical formation of human axillary odor. J Invest Dermatol. 2010 Feb;130(2):529-40. doi: 10.1038/jid.2009.254.

Supplementary material

Table 1. ClinVar's Disease Name and a significance of corresponding SNP found in a sample.

Amyotrophic_lateral_sclerosis_type_10.TARDBP.related_frontotemporal_dementia	Uncertain_significance
not_provided	Benign
Inborn_genetic_diseases	Uncertain_significance
Inborn_genetic_diseases	Uncertain_significance
not_provided	Benign
Cardiovascular_phenotype	Likely_benign
Generalized_epilepsy_with_febrile_seizures_plus,_type_7|Neuropathy,_hereditary_sensory_and_autonomic,_type_2A	Uncertain_significance
Dilated_cardiomyopathy_1G|Autosomal_recessive_limb-girdle_muscular_dystrophy_type_2J	Uncertain_significance
Inborn_genetic_diseases	Likely_benign
not_provided|Familial_adenomatous_polyposis_1|Hereditary_cancer-predisposing_syndrome	Uncertain_significance
Inborn_genetic_diseases	Uncertain_significance
not_provided	Likely_benign
not_provided	Uncertain_significance
Cardiovascular_phenotype	Likely_benign
Inborn_genetic_diseases	Uncertain_significance
not_provided	Likely_benign
not_provided|Early_infantile_epileptic_encephalopathy_with_suppression_bursts	Likely_benign
Telangiectasia,_hereditary_hemorrhagic,_type_2	Likely_benign
Retinoblastoma	Likely_benign
Spastic_paraplegia_52,_autosomal_recessive	Uncertain_significance
Inborn_genetic_diseases	Uncertain_significance
Congenital_myasthenic_syndrome_4A	Uncertain_significance
Inborn_genetic_diseases	Uncertain_significance
Neurofibromatosis,_type_1|Cardiovascular_phenotype|Hereditary_cancer-predisposing_syndrome	Uncertain_significance

Malignant_tumor_of_prostate	Uncertain_significance
not_provided	Uncertain_significance
Rhabdoid_tumor_predisposition_syndrome_2	Likely_benign
Developmental_and_epileptic_encephalopathy,_30	Likely_benign