Practice

BIOINFORMATICS INSTITUTE

Building a perfect human

Valeria Lobanova[1] and Valeria Rubinova[1]

[1]Bioinformatics Institute

Abstract

Microarray genotyping is becoming more and more widely used in genetic laboratory diagnostics, due to the low cost of research and the convenience in choosing a panel of genes based on a particular case. In this work, we used raw 23andMe data with the information about 700k known SNPs. We tried to determine the sex of the patient, the color of his eyes and skin from the information about SNPs. We also proposed to make changes to some SNPs using the technology CRISPR/Cas9 and explained the mechanism of action and the biological value of the changes.

Keywords: SNP, Genotyping chips; CRISPR/Cas9

Introduction

DNA biochip technology is a valuable tool for detecting genetic variation in both research projects and routine screening and has become popular in genetic laboratory research in recent years. SNP chips are DNA microarrays that test genetic variation at many hundreds of thousands of specific locations across the genome. They were initially designed for testing single nucleotide polymorphisms (SNPs) that are common in the population (>1 in 100 people) Bumgarner (2013). SNP chips have proven to be excellent for studying common genetic variation, which can be used to assess ancestry, as well as predisposition to many complex multifactorial diseases such as type 2 diabetes. The genetics community generally recognises that SNP chips perform poorly for genotyping rare genetic variants Weedon *et al.* (2021).

CRISPR/Cas9 – a specific, efficient and versatile gene-editing technology we can harness to modify, delete or correct precise regions of our DNA. CRISPR/Cas9 edits genes by precisely cutting DNA and then letting natural DNA repair processes to take over. The system consists of two parts: the Cas9 enzyme and a guide RNA Redman *et al.* (2016). This technology allows one to precisely manipulate virtually any genomic sequence specified by a short stretch of guide RNA, allowing elucidation of gene function involved in disease development and progressions, correction of disease-causing mutations, and inactivation of activated oncogenes or activation of deactivated cancer suppressor genes when utilizing a fusion protein of nuclease-deficient Cas9 and effector domain Jiang and Doudna (2017).

In this work, we analyzed a dataset with a collection of SNPs obtained using a genotyping chips. After researching a set of some SNPs, we predicted some human phenotypic manifestations and made several substitutions in his genome using the CRISPR-Cas9.

Materials and methods

We analysed the raw 23andMe SNP personal dataset from FixMyProfessor.com. This data was getting in May 2017, so we used the GRCh37 build of human genome as reference.

At first we ran Plink 1.90 beta 7 (Chang *et al.* (2015)) to recode 23andMe raw data to VCF format (with parameters –output-chr MT –snps-only just-acgt). Next we identified haplogroup on mitochondrial chromosome with MThap (Van Oven and Kayser (2009)) and HaploGrep 2.0 (Weissensteiner *et al.* (2016)). Then we defined SNPs that associated with eye and skin colors by 8-plex system from Hart *et al.* (2013). Also we annotate SNPs with Variant Effect Predictor (VEP) (McLaren *et al.* (2016)). Visualization of SNP positions was carried out with Integrative Genomics Viewer (Thorvaldsdóttir *et al.* (2013)). We did sex annotation and filtered the output data from VEP with bash commands and python script. Allele analysis was completed manually with open data from SNPedia (Cariaso and Lennon (2012)).

Results

We defined the haplogroup of research data as H2a2a1 (best match, without mismatches).

Figure 1 Scheme of determining haplogroup by SNPs with HaploGrep 2.0.

In our data we found Y chromosome annotations and absolute majority of X chromosome SNP was homozygotic, so we concluded that the sex of our subject is male.

We extract alleles of key SNPs for 8-plex method to define eye and skin color (Table 1).

This method based on combination of SNPs alleles. However, it doesn't cover all possible combinations of nucleotides, so in some cases, especially if there are heterozygotic allels, it is not allowed to be an accurate result prediction. For eye color we defined our prediction as "not blue" eyes, and skin color wasn't predicted at all.

Table 1 SNP used for determining eye and skin color

Chr	Position	rsID	Ref	Alt	Allele
5	33951693	rs16891982	C	G	0/1
6	341321	rs12203592	C	T	0/1
14	92773663	rs12896399	G	.	0/0
15	28365618	rs12913832	A	G	0/1
15	48426484	rs1426654	A	.	0/0
16	89986154	rs885479	G	.	0/0
Not identified as SNP					
15	27942626	rs1545397			A/A
20	34197406	rs6119471			C/C

We selected some SNPs to be corrected with CRISPR-Cas9 system to improve health on genetic level and determine eye and skin color more definitely. We filtered all the SNPs annotated as potential risk factors and chose those of them, the change in the alleles of which can significantly affect the quality of life and health, according to SNPedia.

At all we suggest 10 SNPs to correct them with DNA editing systems (Table 2).

Table 2 SNP suggested for CRISRP-Cas9 correction.

rsID	Original alleles	Corrected alleles	Gene
rs12913832	G/A	A/A	HERC2
rs6119471	C/G	G/G	ASIP
rs4444903	A/G	A/A	EGF promoter
rs2004640	G/T	G/G	IRF5
rs1801394	G/A	A/A	ASTKD3, MTRR
rs5174	C/T	C/C	LRP8
rs1801274	A/G	G/G	FCGR2A
rs2073658	C/T	C/C	TSTD1
rs699	A/G	A/A	AGT
rs6265	C/T	C/C	BDNF

Discussion

Mechanisms of action for suggested changes:

rs12913832 GA → AA: SNP is located in the gene HERC2. As there is also GG at rs12896399 in the researched genome, the person will have brown eye color after changes Hart *et al.* (2013).

After these changes: rs6119471 CG → GG the person will have not light skin color. This may be important in reducing the risk of developing melanoma, as light skinned people are more susceptible to UV radiation. Roider and Fisher (2016)

We can reduce the risk of glioma and liver cancer in cirrhotic patients by modifying the SNP rs4444903 AG → AA. This SNP, also known as +61, is located in the promoter region of the epidermal growth factor EGF gene that influences the amount of EGF produced. The rs4444903(G) allele appears to produce higher amounts of EGF than the (A) allele Tanabe *et al.* (2008).

rs2004640, a SNP in the IRF5 gene in chromosomal region 7q32.1, is one of several SNPs associated with systemic lupus erythematosus (SLE). So changes made to this SNP GT → GG can decrease the risk of developing systemic lupus erythematosus (SLE) Sigurdsson *et al.* (2008).

It is possible to reduce the risk of developing meningioma by making changes to the SNP rs1801394 GA → AA (genes ASTKD3, MTRR). G/G genotypes associated with increased 5,10-methylenetetrahydrofolate levels are associated with elevated risk for these types of brain cancer Bethke *et al.* (2008).

Change SNP rs5174 CT → CC reduces the risk of heart disease. rs5174 encodes a variant of the LRP8 gene, encoding the low density lipoprotein receptor-related 8 protein (or the apolipoprotein e receptor). The variant affects the protein, changing an arginine to a glutamine Shen *et al.* (2007).

Reduce the risk of developing cancer by editing rs1801274 AG → GG. rs1801274 is a SNP in the Fc fragment of IgG, low affinity IIa, receptor (CD32) FCGR2A gene. rs1801274(C) encodes the arginine (R) allele, with the (T) allele encoding the variant histidine (H). The (H) isoform is considered high-binding to IgG2 and IgG3, while the (R) isoform is considered low-binding Clark *et al.* (1989).

Change rs2073658 CT → CC decreases the risk of developing cardiovascular diseases. rs2073658 a borderline association with metabolic syndrome was observed (p = 0.036, IDF), the minor allele being the risk-increasing allele. The minor allele of rs2073658 is also associated with higher total and LDL-cholesterol, apolipoprotein B-100 and lipoprotein(a) concentrations in longitudinal analyses Auro *et al.* (2008).

rs699 is a SNP in the angiotensin AGT gene that encodes a functional change. So we can reduce the risk of developing hypertension by editing rs699 AG → AA Jeunemaitre *et al.* (1992).

It is possible to reduce the risk of developing depression, ADHD and mental disorders by making changes in rs6265: CT → CC. rs6265, also known as Val66Met, is a SNP in brain-derived neurotrophic factor BDNF gene Krishnan *et al.* (2007).

Literature cited

Auro K, Kristiansson K, Zethelius B, Berne C, Lannfelt L, Taskinen MR, Jauhiainen M, Perola M, Peltonen L, Syvänen AC. 2008. Usf1 gene variants contribute to metabolic traits in men in a longitudinal 32-year follow-up study. Diabetologia. 51:464–472.

Bethke L, Webb E, Murray A, Schoemaker M, Feychting M, Lonn S, Ahlbom A, Malmer B, Henriksson R, Auvinen A *et al*. 2008. Functional polymorphisms in folate metabolism genes influence the risk of meningioma and glioma. Cancer Epidemiology Biomarkers & Prevention. 17:1195–1202.

Bumgarner R. 2013. Overview of dna microarrays: types, applications, and their future. Current protocols in molecular biology. 101:22–1.

Cariaso M, Lennon G. 2012. Snpedia: a wiki supporting personal genome annotation, interpretation and analysis. Nucleic acids research. 40:D1308–D1312.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation plink: rising to the challenge of larger and richer datasets. Gigascience. 4:s13742–015.

Clark M, Clarkson S, Ory P, Stollman N, Goldstein I. 1989. Molecular basis for a polymorphism involving fc receptor ii on

human monocytes. Journal of immunology (Baltimore, Md.: 1950). 143:1731–1734.

Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E. 2013. Improved eye-and skin-color prediction based on 8 snps. Croatian medical journal. 54:248–256.

Jeunemaitre X, Soubrier F, Kotelevtsev YV, Lifton RP, Williams CS, Charru A, Hunt SC, Hopkins PN, Williams RR, Lalouel JM *et al*. 1992. Molecular basis of human hypertension: role of angiotensinogen. Cell. 71:169–180.

Jiang F, Doudna JA. 2017. Crispr–cas9 structures and mechanisms. Annual review of biophysics. 46:505–529.

Krishnan V, Han MH, Graham DL, Berton O, Renthal W, Russo SJ, LaPlant Q, Graham A, Lutter M, Lagace DC *et al*. 2007. Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. Cell. 131:391–404.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The ensembl variant effect predictor. Genome biology. 17:1–14.

Redman M, King A, Watson C, King D. 2016. What is crispr/cas9? Archives of Disease in Childhood-Education and Practice. 101:213–215.

Roider EM, Fisher DE. 2016. Red hair, light skin, and uv-independent risk for melanoma development in humans. JAMA dermatology. 152:751–753.

Shen GQ, Li L, Girelli D, Seidelmann SB, Rao S, Fan C, Park JE, Xi Q, Li J, Hu Y *et al*. 2007. An lrp8 variant is associated with familial and premature coronary artery disease and myocardial infarction. The American Journal of Human Genetics. 81:780–791.

Sigurdsson S, Göring HH, Kristjansdottir G, Milani L, Nordmark G, Sandling JK, Eloranta ML, Feng D, Sangster-Guity N, Gunnarsson I *et al*. 2008. Comprehensive evaluation of the genetic variants of interferon regulatory factor 5 (irf5) reveals a novel 5 bp length polymorphism as strong risk factor for systemic lupus erythematosus. Human molecular genetics. 17:872–881.

Tanabe KK, Lemoine A, Finkelstein DM, Kawasaki H, Fujii T, Chung RT, Lauwers GY, Kulu Y, Muzikansky A, Kuruppu D *et al*. 2008. Epidermal growth factor gene functional polymorphism and the risk of hepatocellular carcinoma in patients with cirrhosis. Jama. 299:53–60.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (igv): high-performance genomics data visualization and exploration. Briefings in bioinformatics. 14:178–192.

Van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial dna variation. Human mutation. 30:E386–E394.

Weedon M, Jackson L, Harrison J, Ruth K, Tyrrell J, Hattersley A, Wright C. 2021. Use of snp chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation. bmj. 372.

Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, Kronenberg F, Salas A, Schönherr S. 2016. Haplogrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic acids research. 44:W58–W63.