Validation of candidate SNPs for human hair morphology using Ion PGM™ NGS technology – towards further extension of forensic DNA phenotyping tools

Ewelina Pośpiech[1,2], Magdalena Kukla[3], Małgorzata Skowron[4], Wojciech Branicki[2]

1 – Institute of Zoology, Faculty of Biology and Earth Sciences, Jagiellonian University, Kraków, Poland

2 – Malopolska Centre of Biotechnology, Jagiellonian University, Kraków, Poland

3 – Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Kraków, Poland

4 – Department of Dermatology, Collegium Medicum of the Jagiellonian University, Kraków, Poland

Introduction

The worldwide variation of human hair morphology is significant and particularly differentiating in Europe. Therefore, DNA-based prediction of this trait has great potential for intelligence purposes in forensics. To date, the strongest association with hair morphology in Europeans has been reported for the trichohyalin gene (*TCHH*) explaining ~6% of the total variance. Weaker effects have been suggested for *WNT10A* and *FRAS1*. Our previous study has shown that analysis of three SNPs in these three genes can predict straight hair with high sensitivity but low specificity. The genotype TT (rs11803731) - GG (rs7349332) - GG (rs1268789) was selected as the best predictor of straight hair giving >80% of probability for straight hair occurrence [Pośpiech et al., 2015]. To further improve forensic DNA-based prediction of human hair morphology we decided to validate extended set of candidate SNPs using high-throughput DNA sequencing technology.

Materials and Methods

Multiple candidate SNPs were selected from the literature based on their potential role in human hair morphogenesis and growth, expression pattern in the hair follicle, involvement in protein-protein interactions and relationship with pathological condition of human hair structure. This extended set of 96 candidate SNPs was analyzed using Ion AmpliSeq™ technology and Ion PGM™ system. DNA libraries were prepared using Ion AmpliSeq Library Kits 2.0 with 5 ng of input DNA and 10 µl of PCR volume reaction per primer pool. DNA libraries were normalized to 100 pM. Libraries for 40 samples were combined in equal ratios and subjected to the template preparation reaction using Ion PGM™ Hi-Q™ OT2 Kit with increased volume of combined libraries from 2 µl to 5 µl comparing to the original protocol. Sequencing was performed with Ion PGM Hi-Q Sequencing Kit with Ion 318 Chip type with one run per initialization and 1000 number of flows. Raw data were analyzed with Torrent Server and variants were typed with Torrent Variant Caller plugin. The results of genotyping were validated using data for 6 SNPs previously associated with hair morphology and typed with SNaPshot technology. The preliminary study involved 82 individuals of curly hair and 63 individuals of wavy hair from Poland.

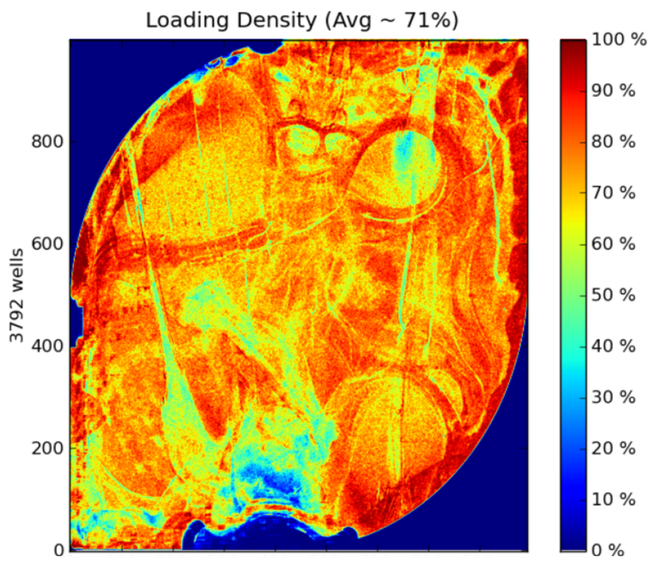

Fig. 1. Chip Ion Sphere Particles loading density for the exemplary sequencing run.

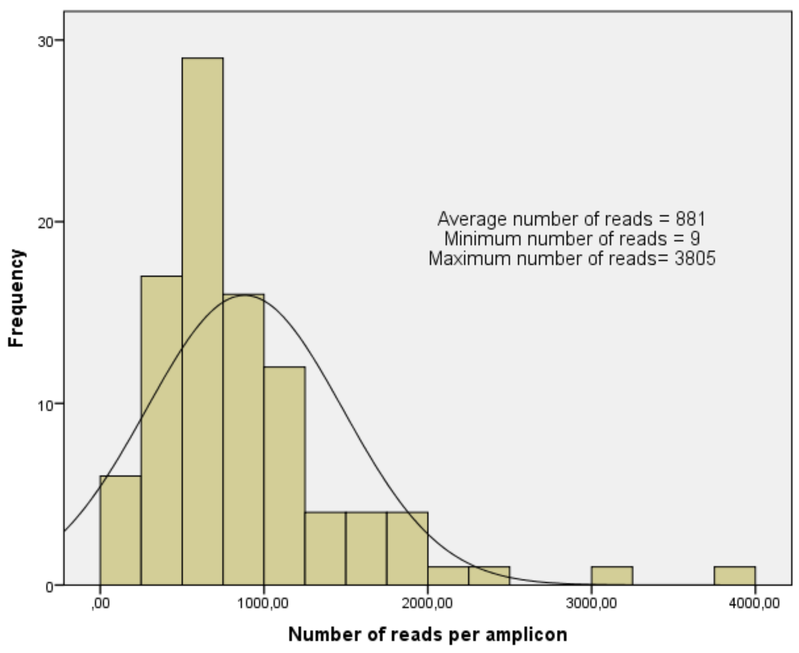

Fig. 2. Distribution of the number of reads per amplicon for the exemplary sample.

Results

Our Ion AmpliSeq™ panel is the size of 28.54 kb and comprised of two pools (51 and 45 SNPs) covering 96 SNPs including most promising candidates for hair morphology. The vast majority of selected SNPs are located on chr1 (36 SNPs), chr2 (16 SNPs) and chr12 (14 SNPs). The size of amplicons is in range 125 – 375 bp with the average length of ~300 bp. The average value of Ion Sphere Particles loading on chip is 66% (Fig. 1) with the increased volume of combined libraries used for template preparation reaction comparing to the original protocol. The average number of total reads per run is 4.3 mln with the average reads per amplicon of ~880 (Fig.2). The percentage of amplicons with at least 1 read is 100% and 97.92% with at least 20 reads (Fig.3.). In details, there are two problematic SNPs with the number of reads between 1-30 (~70% of samples with <10 reads). Concordance study performed for 6 most important hair morphology SNPs with SNaPshot technology provided 100% of consistent results. Among candidate genes selected in the present study is *IGFBP5* (chr2). Nine SNPs within *IGFBP5* were selected based on their MAF in Europe and considering the highest divergency between Europe, Asia and Africa (Table 1). MAF designated for the tested sample are in line with 1000 Genomes database and future studies should validate their association with hair morphology variation.

Amplicon Read Coverage	
Number of amplicons	96
Percent assigned amplicon reads	79.08%
Average reads per amplicon	881.4
Uniformity of amplicon coverage	94.79%
Amplicons with at least 1 read	100.00%
Amplicons with at least 20 reads	97.92%
Amplicons with at least 100 reads	96.88%
Amplicons with at least 500 reads	76.04%
Amplicons with no strand bias	98.96%
Amplicons reading end-to-end	4.17%

Fig. 3. Amplicon read coverage statistics for the exemplary sample.

SNP ID	Gene	Chromosome position (hg19)	MAF in Europe	MAF in Asia	MAF in Africa	MAF in the studied sample
rs4674107	*IGFBP5*	chr2:217536354	C 0.30	C 0.36	C 0.02	C 0.32
rs11575194	*IGFBP5*	chr2:217543728	A 0.03	A 0.00	A 0.00	A 0.07
rs11575161	*IGFBP5*	chr2:217549963	T 0.44	T 0.18	T 0.10	T 0.42
rs741384	*IGFBP5*	chr2:217551954	C 0.50	C 0.40	C 0.45	C 0.47
rs4480966	*IGFBP5*	chr2:217560985	C 0.13	C 0.00	C 0.04	C 0.12
rs1978346	*IGFBP5*	chr2:217561467	G 0.37	G 0.50	A 0.23	G 0.34
rs67587000	*IGFBP5*	chr2:217561648:50	del 0.15	del 0.33	ATG 0.36	del 0.12
rs13015993	*IGFBP5*	chr2:217625523	G 0.28	G 0.47	A 0.48	G 0.24
rs4442975	*IGFBP5*	chr2:217920769	T 0.49	G 0.10	T 0.31	T 0.49

Table 1. Characteristics of SNPs in *IGFBP5* under study.

Discussion

High-throughput technologies of DNA sequencing have revolutionised biomedical and forensics studies and allowed for significant expansion of research on human genetic variation. In the present study we selected multiple candidate SNPs for human hair morphology which is a highly conspicuous and variable trait of human appearance and therefore is of great interest in the field of Forensic DNA Phenotyping. This extended set of candidate SNPs can be effectively analysed only using high-throughput methods like this provided by Ion AmpliSeq™ technology and Ion PGM™ system. Our panel comprises two primer pools and covers 96 selected SNPs. It allows for multiplexing of 40 samples per one sequencing reaction and enables accurate assignment of 94 from 96 tested SNPs. *IGFBP5* gene is a new candidate for human hair morphology variation. It encodes a binding protein for insulin-like growth factor-1 (IGF-1) which has a crucial role in hair development and growth. It has been demonstrated that *IGFBP5* is significantly higher expressed in curly hairs compared to the straight ones [Sriwiriyanont et al., 2011]. Therefore, analysis of polymorphisms relevant to the regulation of expression of this gene, particularly in the promoter region seems to be very promising in the context of genetic determination of hair morphology in humans. Extended studies should verify its utility for human hair prediction for forensic purposes.

This research was supported by the grant from the European Union Seventh Framework Programme under grant agreement no. 285487 (EUROFORGEN-NoE) and a grant from the National Science Centre in Poland no 2014/15/D/NZ8/00282.