ForenSeq® Imagen Kit

Forensic community-approved SNPs for appearance and biogeographic ancestry estimations

Highlights

- ***Options to support regional regulations***
  Evaluate phenotypic SNPs with or without biogeographical ancestry.
- ***High sensitivity for challenging samples***
  Acquire investigative lead generation data with as little as 32 pg of DNA.
- ***Interoperable with HIrisPlex-S software***
  Interpret NGS data with established tools like the HIrisPlex-S interactive website.

Introduction

Recent advances in forensic genetics have enabled the analysis of new markers that help with forensic investigations. While short tandem repeats (STRs) are still the predominant forensic genetic marker for identity testing and kinship analysis, single nucleotide polymorphisms (SNPs) have emerged as a powerful tool for investigative lead generation (1). SNPs are point mutations that encompass single-base substitutions and single-base insertions and/or deletions (InDel) and occur ubiquitously in coding and non-coding regions of the genome. SNPs represent the most common human genetic variation, occurring once every 1000 bp. SNPs have a lower mutation rate and lend themselves to smaller amplicon sizes of 50–150 bp. This makes SNPs particularly suitable in cases of aged, degraded or low copy biological samples; in long-range kinship and paternity testing; and in population and evolutionary genetics research (2).

Most SNPs are bi-allelic markers, and consequently, in the diploid human genome, there are only three possible genotypes (AA, BB or AB) associated with each SNP. This lower discrimination power per SNP means that numerous loci must be tested to yield the same discriminative power as STRs (3). Capillary-Electrophoresis (CE)-based methods struggle to support an expanded number of markers but low cost next-generation sequencing (NGS) removes this limitation and enables the use of SNP-based forensic panels for differentiating individuals from one another. Common applications for SNPs include kinship/paternity testing and evolutionary studies, biogeographical ancestry (BGA) and estimation of visible traits or appearance, such as skin, hair or eye color, height, weight,

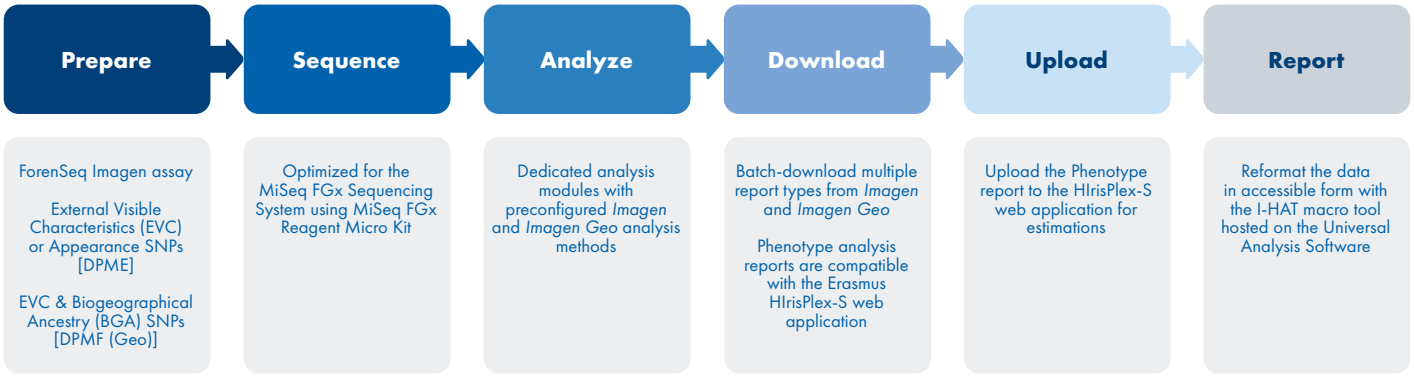

Figure 1.
ForenSeq Imagen Kit workflow.

facial morphology etc., commonly known as External Visible Characteristics (EVCs) (3, 4, 5 ,6). The analysis of EVC and BGA could be especially useful in investigative cases where there are no potential suspects and no match between the evidence DNA sample under investigation and genetic profiles entered into criminal databases.

The ForenSeq Imagen workflow, comprising ForenSeq Imagen Kit, the MiSeq FGx® Sequencing System with MiSeq FGx Reagent Micro Kit and the ForenSeq Imagen analysis methods in Universal Analysis Software (UAS), is a fully validated DNA-to-data workflow specifically designed for forensic genomics applications (Figure 1).

ForenSeq Imagen relies on established, peer-reviewed SNP marker sets that have been identified and tested by the forensic community. The kit is an optimized solution for operational forensic labs considering implementing an NGS-based lead generation workflow (4, 6, 7).

Forensically optimized SNP multiplex modularized for regional regulatory compliance

ForenSeq Imagen is a fully kitted solution that includes all the reagents and amplification controls required for the generation of 96 unique DNA libraries for sequencing on the MiSeq FGx Sequencing system using MiSeq FGx Reagent Micro Kit. To ensure compatibility with existing and emerging regulations, two primer pools, one comprised of markers for EVC, also referred to as DPME, and another comprised of markers for a combination of EVC and BGA, also referred to as DPMF (Geo). Both primer sets encompass 41 SNPs for the phenotypic estimations that include 22 SNPs for the estimation of hair color, 6 SNPs for the estimation of eye color, 36 markers for the estimation of skin color, with some SNPs common to more than one phenotype. Both primer sets encompass 14 Y-SNPs to enable robust estimation of biological sex. In addition, 56 SNPs for the estimation of biogeographic ancestry are included in one of the primer sets. Labs have the flexibility to use either of the two primer sets based on the regulatory approaches associated with their geography. Table 1 provides a summary of the number of markers associated with both primer sets in ForenSeq Imagen Kit (8).

ForenSeq Imagen Kit uses established ForenSeq chemistry (9). ForenSeq Imagen is a six-step method that leads to the generation of high-quality sequencing libraries in 7 hours and 15 minutes, with just 1 hour and 30 minutes of hands-on time. The kit is optimized for the most common sources of forensic DNA, utilizing 8 µl of DNA extract input volume. This sensitive, PCR-based assay is designed to generate short amplicons and is integrated with Unique Dual Indices (UDIs), maximizing the likelihood of generating DNA profiles with degraded and low-input samples. Either the kitted buffer or the optional ForenSeq Enhanced PCR1 Buffer System can be used for challenging samples to enable support for inhibited forensic samples. The workflow includes five safe stopping points and a premixed adapter plate that increases library preparation efficiency and ease.

Table 1. Categories and number of SNP markers in ForenSeq Imagen Kit

SNP category	Number of markers	Amplicon size range (bp)	Included in DNA primer	
			Mix E	Mix F
Y-SNPs	14	88–130	Yes	Yes
Phenotypic SNPs*	41	69–227	Yes	Yes
Hair color	22	92–227	Yes	Yes
Eye color	6	73–119	Yes	Yes
Skin color	36	69–213	Yes	Yes
Biogeographical† ancestry SNPs	56	67–200	No	Yes

*Four SNPs overlap across eye, skin and hair color phenotypes; 13 SNPs overlap across hair and skin color phenotypes; and 2 SNPs overlap across eye and skin color phenotypes.
† Only available in the EVC+BGA SNP panel [DPMF (Geo)].

High multiplexing capability maximizes sample throughput while minimizing cost per sample. A low gDNA input recommendation of 1 ng enables reliable and reproducible recovery of full profiles from high-quality single-source samples all the way down to 32 pg low input and difficult samples using the EVC+BGA [DPMF (Geo)] primer set and 16 pg for the EVC (DPME) primer set. ForenSeq Imagen enables easy and automatable access to an operational workflow for EVC and BGA. Table 2 provides a complete list of kit specifications.

Table 2. ForenSeq Imagen Kit specifications

Specification	Value
Sample type	Extracted gDNA from a range of forensically relevant sample types including punches from storage cards (FTA®)
Recommended input	1 ng per sample (gDNA) 1.2 mm per samples (FTA card punch)
Kit configuration	96 reactions
Primer sets	2
Marker composition	DPME: 41 EVC markers, 14 Y-SNPs DPMF (Geo): 41 EVC, 56 BGA markers, 14 Y-SNPs
Amplicon size (mean)	DPME - 112 bp DPMF (Geo) – 111 bp
Recommended sequencing plexity	8 to 96 samples on MiSeq FGx Reagent Micro Kit
Library prep time	7 hours 15 minutes (total) 1 hour 30 minutes (hands-on)

Robust inhibition tolerance and high sensitivity enable accurate SNP detection

Forensic DNA samples are frequently challenged by a variety of environmental factors and PCR inhibitors such as hematin, humic acid, tannic acid and indigo. The presence of these inhibitors reduces the amplification efficiency of the workflow (2, 9, 10). ForenSeq Imagen Kit contains a robust inhibition buffer (PCR1) that enables efficient amplification in the presence of these commonly occurring inhibitors. For severely challenged samples such as interred bones, ForenSeq Imagen Kit is also compatible with the ForenSeq Enhanced PCR1 Buffer System (ePCR1). ForenSeq Imagen demonstrates high call rates of 100% with 0.63 ng/µL of humic acid, 1 µM of tannic acid and 133 µM of Indigo, and 97% with 20 µM of Hematin, when processed with the DPMF (Geo) primer mix. The enhanced buffer system demonstrates a 20X higher tolerance for humic acid, 6X higher tolerance for hematin and 4X higher tolerance for tannic acid compared to PCR1 buffer. Figure 2 summarizes studies on inhibition tolerance with DPMF (Geo). Inhibition studies with DPME show similar results.

The ability to generate accurate SNP calls across a range of DNA inputs was evaluated using serially diluted gDNA in template amounts of 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, 31.25 pg, 15.625 pg and 7.82 pg, in triplicate. A total of 96 libraries were

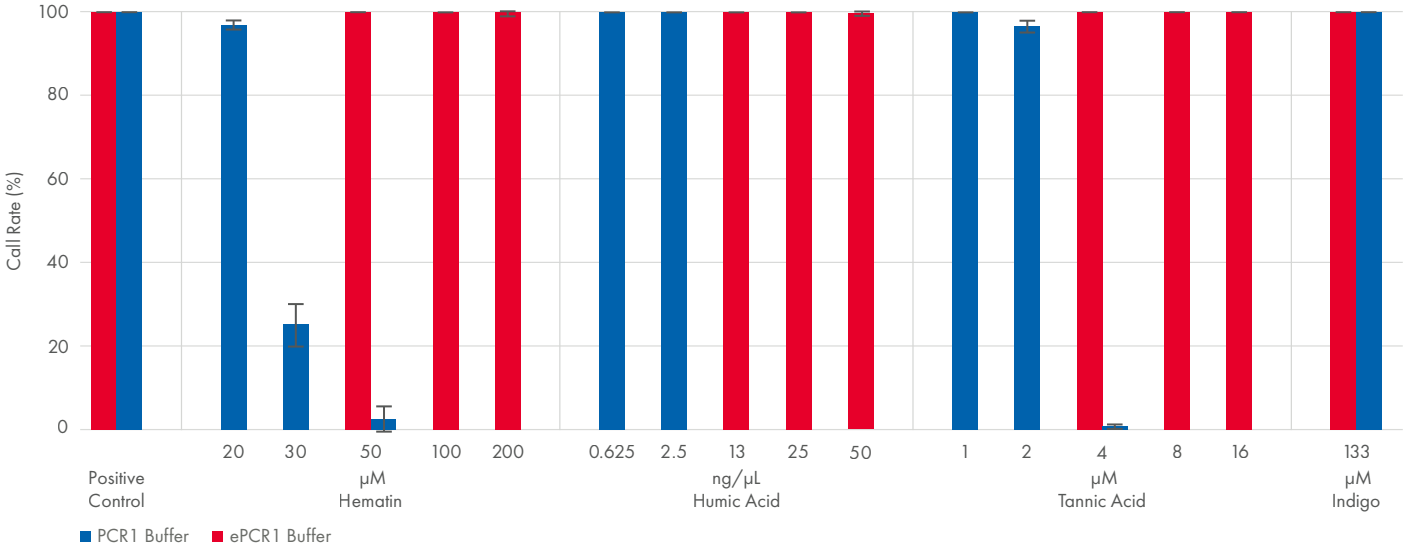

Figure 2.
Inhibition Study. Samples gDNA (1 ng) treated with common forensic inhibitors, processed with both PCR1 and ePCR1 buffers and the DPMF (Geo) primer mixes in triplicate, show high average call rates.

simultaneously sequenced using MiSeq FGx Reagent Micro Kit. ForenSeq Imagen offers a high level of sensitivity, generating 100% call rate with as little as 32 pg of input DNA with both primer sets and 94% call rate with 16 pg of input DNA using DPMF (Geo), when 96 libraries were simultaneously sequenced. Figure 3 summarizes the sensitivity studies.

Reproducible high-quality results with mock casework and degraded samples

ForenSeq Imagen Kit enables generation of high-quality SNP profiles from mock casework and degraded samples. Evaluation of profiles from 1 ng of DNA extracted from saliva-FTA cards, blood, buccal swabs, plucked hair, teeth and bones subject to a variety of insults demonstrated an average call rate of 100%. Similar studies were performed on 1 ng of DNA extracted from a blood sample subjected to artificial degradation to generate a degradation series spanning a degradation index (DI) of 1 to 50. Samples processed with the DPME primer mix showed an average call rate of 100% up to a DI of 26 and a 98% call rate for a DI of 50. The DPMF (Geo) primer mix showed a call rate of 100% up to a DI of 8 and a 98% call rate for a DI of 50. ForenSeq Imagen Kit enables forensic laboratories to generate high quality SNP markers across a range of sample sources and degradation levels. Table 3 and Figure 4 summarize studies with mock case work and degraded samples, respectively (11).

ForenSeq Imagen Kit uses forensic community-approved SNP marker sets for the estimation of phenotypes and biogeographic ancestry. These established markers have been previously tested by multiple academic and operational forensic laboratories. To supplement these studies, 60 samples with known genotypes were sequenced with both DPME and DPMF (Geo) as part of 96 sample runs. Data from the 1000 Genomes Project served as the truth set for concordance and accuracy studies (12). A high concordance rate of 99% and accuracy rate of 99% was observed with more multiplexes. Precision was calculated using data from 121 positive controls generated by 10 users. High precision rates of 99% were observed with both primer multiplexes. In addition, accuracy of EVC or appearance predictions were made by the HIrisPlex-S web application using SNP profiles generated for 1 ng of 11 reference samples that were processed using DPMF (Geo) and sequenced in triplicate by three operators (5).

Data from the 1000 Genomes Project was used to evaluate the accuracy of predictions. All samples demonstrated 100% call rates and 100% accuracy of predicted phenotypes (Table 4). Accuracy of ancestry predictions made by the UAS was evaluated using DPMF (Geo) SNP profiles generated from 1 ng DNA across 58 samples that were part of the 1000 Genomes Project. All samples demonstrated 100% call rates and 100% accuracy of ancestry predictions across four population groups of Caucasian, African, admixed American, and Asian populations.

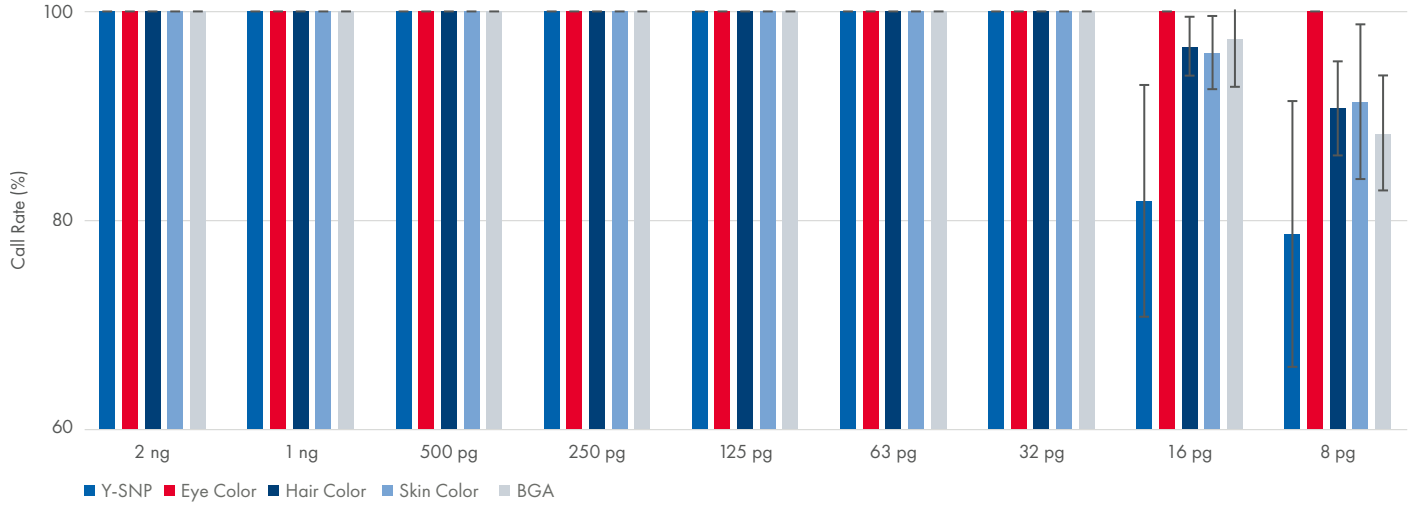

Figure 3.
Sensitivity Study. Serially diluted gDNA samples processed with the DPMF (Geo) primer mix in triplicate show high average call rates.

Table 3. Mock casework studies

Sample	Treatment/Collection method	Average call rate (%)	
		DPMF	DPMF (Geo)
Blood	No heme	100	100
	Light heme	100	100
	Moderate heme	100	100
	Heavy heme	100	100
Bone	Burned (2 samples)	100	100
	Cremated (2 samples)	100	100
	Embalmed (2 samples)	100	100
Saliva	FTA card	100	100
Buccal	Swab	100	100
Hair	Plucked	100	99.50
Teeth		100	99.80

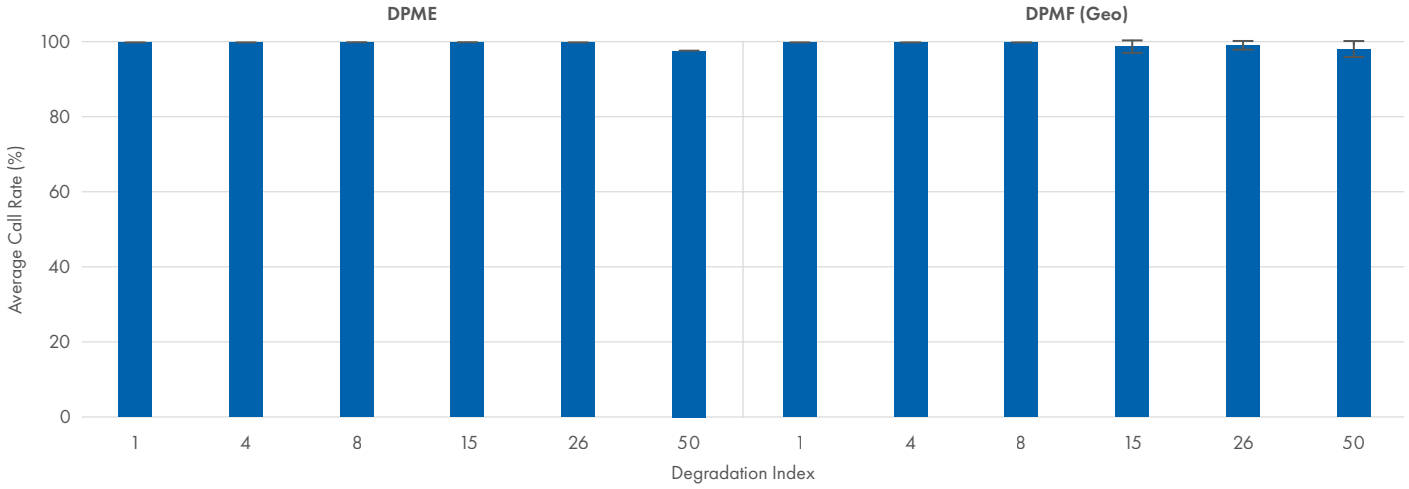

Figure 4.
Degradation Study. Serial degradation of 1 ng gDNA samples processed with DPMF (Geo) primer mix in triplicate show high average call rates.

End-to-end workflow integration across analysis and interpretation

Data generated by the ForenSeq Imagen kit is optimized for analysis using the Universal Analysis Software. Two fit-for-purpose analysis methods enable data analysis from the DPME or DPMF (Geo) primer mixes separately. Robust preconfigured thresholds and settings allow easy validation and implementation. Rich data visualization capabilities enable labs to quickly gauge the quality of their sequencing run and SNP calls. Extensive filtering and sorting capabilities streamline manual review of markers that require inspection. The UAS generates SNP reports that can be uploaded to the HIrisPlex-S web application for predictions powered by solutions described in multiple peer-reviewed publications (13). To allow easy interpretation of complex statistical information, reports from the HIrisPlex-S web application (14) can be parsed by

Table 4. ForenSeq Imagen generates highly concordant genotypes when compared with orthogonal technologies, as well as high precision and accuracy rates with both DPME and DPMF (Geo)

	DPME	DPMF (Geo)
Concordance rate (60 samples)	99.8%	99.8%
Accuracy (121 samples)	99.8%	99.8%
Precision (58 samples)	99.9%	100%

the Imagen-HIrisPlex-S Analysis Tool (I-HAT) application on the UAS data server. The I-HAT macro generates human-readable reports that summarize the predictions and probabilities associated with typed SNPs. Automation of the lab workflow allows labs to onboard the entire ForenSeq Imagen workflow from lab to report with confidence.

Conclusion

ForenSeq Imagen offers an easy solution for the inclusion of NGS-based investigative intelligence into operational workflows. ForenSeq Imagen Kit contains reagents to sequence SNP markers that estimate hair, eye, skin color and biogeographic ancestry. Modular components ensure that labs can analyze only the markers that are approved by their regional regulatory body. Low input recommendations of 1 ng gDNA deliver a reproducibly high call rate regardless of inhibition or degradation. Seamless integration of the data with the Imagen Analysis Module across the HIrisPlex-S web-application and I-HAT macro tool allows laboratories to easily sequence, analyze, interpret and report data in less than 1.5 days.

Ordering Information

Product	Contents	Cat. no.
ForenSeq Imagen Kit	Includes the required reagents to process up to 96 samples simultaneously for phenotypic and biogeographical ancestry information to generate comprehensive DNA profiles	V16000189
MiSeq FGx Reagent Micro Kit	Supports up to 5 million paired-end reads for small batch sizes and faster turnaround times	20021681
MiSeq FGx Sequencing System	Desktop instrument with two run modes for a range of forensic genomics applications within a validated NGS workflow	15048976
ForenSeq Universal Analysis Software	Software pre-installed as a dedicated server specific for forensic genomics for run setup, sample management, analysis and report generation. This product includes server, mouse, keyboard and monitor.	9003364

Learn more about NGS for HID in your lab. Visit **qiagen.com/ForenSeqImagen**

References:

1. Sobrino B, Carracedo A. SNP typing in forensic genetics: a review. *Methods Mol Biol.* 2005; 297: 107-26.
2. Kidd KK, et al. Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int.* 2006; 164: 20–32.
3. Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: Thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol.* 2007; 3: 200–205.
4. Sanchez JJ, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification." *Electrophoresis* 2006; 27: 1713–24.
5. Chaitanya L, et al. The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci Int Genet.* 2018; 35: 123–35.
6. Kidd KK, et al. Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Sci Int Genet.* 2014; 10: 23–32.
7. Palencia-Madrid L, et al. Evaluation of the VISAGE Basic tool for appearance and ancestry prediction using PowerSeq chemistry on the MiSeq FGx System. *Genes* 2020; 11: 708.
8. *Verogen ForenSeq Imagen Reference guide* - VD20222008
9. Jäger AC, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Sci Int Genet.* 2017; 28: 52–70.
10. Nakazato T, Ohta T, Bono H. Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. Edited by Ramy K. Aziz. *PLoS ONE* 2013; 8: e77910.
11. InnoGenomics InnoQuant HY. https://innogenomics.com/products/innoquant-hy/
12. 1000 Genomes Project | Info and Population Samples | Coriell. https://www.coriell.org/1/NHGRI/Collections/1000-Genomes-Project-Collection/1000-Genomes-Project
13. Sharma V, et al. Evaluation of ForenSeq™ Signature Prep Kit B on predicting eye and hair coloration as well as biogeographical ancestry by using Universal Analysis Software (UAS) and available web-tools. *Electrophoresis* 2019; 40: 1353–64.
14. HIrisPlex-S Eye, Hair and Skin Colour DNA Phenotyping Webtool. https://hirisplex.erasmusmc.nl.