

Development of targeted amplicon next-generation sequencing panel of 50 SNPs related to externally visible characteristics and behavior

Hee-Yeon Park, Yoonji Noh, Eung-Soo Kim[★], and Hyun-Chul Park[★]

Forensic DNA Division, National Forensic Service, 10, Ipchun-ro, Wonju-si, Gangwon-do 26460, Korea

(Received March 13, 2014; Revised April 12, 2024; Accepted May 8, 2024)

외형 및 행동 습관 관련 50개 SNP 마커 분석을 위한 targeted amplicon next-generation sequencing 패널 개발

박희연 · 노윤지 · 김응수[★] · 박현철[★]

국립과학수사연구원 유전자과

(2024. 3. 13. 접수, 2024. 4. 12. 수정, 2024. 5. 8. 승인)

Abstract: In forensic genetics, when it is not possible to confirm an individual's identity through STR profile analysis, additional information about the individual can be obtained using DNA-based phenotypic traits estimation. Recently, various researches have been conducted on methods to determine externally visible characteristics (EVC) such as eyes, hair, and skin color. However, relying solely on such phenotypic traits information has limitations for application in East Asian regions, including Korea. In this study, in order to utilize EVC related to an individual's appearance as investigative information, SNPs related to eye shape, hair thickness, skin color, as well as baldness, body type, high myopia, facial shape, acne, and behavioral habits were explored. A total of 50 SNPs were selected, and a targeted amplicon NGS panel capable of amplifying them all at once was developed. Experimental results confirmed the allelic types and frequencies of the 50 SNPs in 14 samples. We plan to use this panel to investigate the correlation between genotype and phenotype using various samples, and to develop methods for interpreting the results.

요약: 범유전학에서 개인의 신원확인을 위한 STR 프로파일 분석이 불가능한 경우, DNA를 이용한 외형추정특성을 이용하여 개인에 대한 정보를 얻을 수 있다. 최근 눈동자, 머리카락, 피부 색과 같은 외형추정특성을 확인하는 방법들이 연구되고 있지만, 이러한 외형추정특성 정보만 가지고는 한국을 비롯한 동아시아 지역에서 적용하기에는 한계가 있다. 본 연구에서는 개인의 외형과 관련된 표현형을 수사정보로서 활용하기 위해 눈 모양, 머리카락 굵기, 피부 색 뿐만 아니라 탈모, 체형, 고도근시, 얼굴모양, 여드름, 행동습관과 관련된 SNP를 탐색하였다. 이들 표현형과 관련된 50개의 SNP를 선정하여 한 번에 증폭할 수 있는 targeted amplicon NGS 방식의 multiplex PCR 패널을 개발하였다. 실험 결과 14개 샘플에서 50개 SNP의 대립유전자 유형과 빈도를 확인할 수 있었다. 향후 본 패널

[★] Corresponding author

Phone : +82-(0)33-902-5724 Fax : +82-(0)33-902-5946

E-mail : hcpark79@korea.kr

This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

을 가지고 더 많은 샘플을 이용하여 유전형과 표현형 간 연관성 확인 및 결과 해석 방법을 분석할 예정이다.

Key words: SNP, externally visible characteristics, targeted amplicon NGS, multiplex PCR

1. 서 론

법유전학 분야에서 개인 식별은 주로 Short Tandem Repeat (STR)을 이용하여 분석을 수행한다.¹ STR은 특정 DNA의 짧은 단위 서열(2~7 bp)이 반복적으로 나타나는 구조로 되어 있고, 이 단위 서열의 반복 횟수는 사람마다 다르며, 이러한 반복 횟수의 변이를 통해 각각의 사람들은 고유한 DNA 프로필을 보유하게 된다.² 법과학 시료에서 STR 분석으로 확인된 DNA 프로필은 용의자와의 비교 또는 DNA 데이터베이스의 검색을 통하여 사건 해결에 필요한 단서를 제공한다.

이러한 STR은 신원확인을 위해 가장 강력한 도구임이 틀림없지만, 다른 유전정보를 가지고 있지 않기 때문에, 일치하는 대상자가 없는 경우에는 범인을 추정할 수 없을 뿐만 아니라, 불상 변사자의 신원확인에도 있어 어려움이 있다. 이를 보완하기 위해 법유전학 분야에서는 DNA로부터 외부적으로 보이는 특성(Externally visible characteristics, EVC)과 연관된 변이를 찾고, 이를 이용하여 표현형을 예측하고자 하는 연구가 꾸준히 진행되어 왔다.^{3,5} 이러한 EVC에 대한 분석을 통하여 개인의 외형을 예측하는 것은 수사에 대한 정보와 단서를 제공하여 용의자의 대상을 좁혀 나가는데 이용할 수 있다. 그리고 실종 사건이나 재난 사건에서 개인 식별에 대한 정보를 제공할 수 있다.^{5,6}

EVC에서 분석되는 부위는 대부분 Single Nucleotide Polymorphism (SNP)이고, SNP는 DNA 염기의 치환, 삽입, 결실을 말하며 낮은 돌연변이율을 가지고, 유전성이 높은 것으로 알려져 있다.⁷ 또한, 분석 대상 시료에서 작은 크기의 PCR 증폭 산물로도 확인할 수 있어 DNA가 손상된 시료나 적은 양의 농도 DNA 시료에서도 분석이 용이하다.⁸ 특히 SNP 중 눈동자색과 머리카락 색깔은 개인의 표현형을 구별하는데 자주 사용되는 외형 특징이며, 이러한 특성은 유전성이 매우 높고 상대적으로 적은 수의 유전자와 연관되어 있어 표현형 분석에 적합한 것으로 알려져 있다.⁵ SNP를 이용한 눈동자, 머리카락, 피부 색 등과 같은 외형 추정 특성과 관련된 표현형 연구는 법유전학 분야에서 지속적으로 연구되어 왔다. 현재까지의 SNP에 대한

연구 및 외형 예측 시스템의 대부분은 서구 인구의 특성을 대상으로 이루어졌으며, 인구 혼잡이 많은 인구 집단에서의 예측에 적합하다.⁹

기존 상용화된 Next-Generation Sequencing (NGS) 패널은 상염색체 STR (A-STR), 남성 성염색체 STR (Y-STR) 및 다양한 SNP 등 많은 정보를 담고 있지만, 외형 추정 특성 표현형 관련 SNP의 경우 눈동자 색, 머리카락 색, 피부 색에 제한되어 있으며, SNP 유전형을 이용한 표현형 연관 SNP 분석에서도 제한적인 정보만을 제공한다. 눈동자 색, 머리카락 색, 피부 색의 예측은 환경 요인의 영향이 적어 예측 결과의 신뢰도가 높지만, 한국인을 비롯한 동아시아 집단에서는 눈동자 색과 머리카락 색을 이용한 구분에 어려움이 있다.¹⁰

본 연구에서는 눈동자 색, 머리카락 색 외 다른 외형을 특정하기 위해 피부 색과 관련된 SNP 17개, 눈모양, 머리카락 굵기, 피부 색, 탈모, 체형, 고도근시, 얼굴모양, 여드름과 관련된 SNP 28개(외형 추정)와 알코올 대사체 및 니코틴 대사체와 같이 행동 추정과 관련된 SNP 5개를 탐색하였고, 이들 SNP를 한 번에 분석하기 위해 targeted amplicon NGS 방식의 multiplex PCR 패널을 개발하였다. 본 패널을 이용하여 총 14개 샘플에서 50개 SNP가 성공적으로 분석되었고, 각 SNP의 대립유전자 빈도(allele frequency)도 확인할 수 있었다.

법과학 시료에서 외형 예측 방법을 적용하기 위해서는 표현형을 구체적으로 구별할 수 있는 외형 관련 SNP 마커를 찾고, SNP 마커의 분석을 통한 외형 예측 모델을 개발하는 것이 매우 중요하다. 따라서 본 연구에서는 법과학 시료에 적용 가능한 외형 추정 및 행동 습관 관련된 SNP를 선별하고, 선별된 다수의 SNP 마커를 NGS 기반으로 분석하기 위한 패널을 개발하였다.

2. 재료 및 방법

2.1. 시료 채취

본 연구에서 사용한 시료는 국립과학수사연구원 기

관생명윤리심사 승인(IRB: 906-240131-BR-016-01)에 따라 14명의 구강 시료를 채취하여 사용하였다. 14명은 흡연자, 음주 가능 여부, 탈모, 주근깨 여부, 체형 차이 등을 고려하여 선별하였고, 남성과 여성 각각 7명의 시료를 사용하였다.

2.2. DNA 추출 및 정량

본 연구에서는 구강 샘플을 사용하였다. QIAamp DNA Micro Kit(Qiagen, Hilden, Germany)를 이용하여 제조사의 매뉴얼에 따라 DNA를 추출하였다. 추출한 DNA의 최종 부피는 100 μ L가 되도록 하였다. 추출된 DNA는 Quantifiler™ Trio DNA Quantification Kit (Thermo Fisher Scientific, Waltham, MA, USA)의 사용자 매뉴얼에 따라 7500 Real-Time PCR system (Thermo Fisher Scientific, Waltham, MA, USA)을 이용하여 정량분석을 하였다.

2.3. 프라이머 제작

각 SNP 위치의 앞과 뒤 100 bp 서열을 포함한 201 bp 서열을 대상으로 프라이머 부위를 탐색하였다. National Center for Biotechnology (NCBI)의 Primer designing tool 및 Primer3 프로그램을 기반으로 프라이머를 디자인하였다. 각 50개 마커의 증폭 산물 길이가 135 bp에서 150 bp 사이에 위치하도록 하였으며, GC% 범위는 30%에서 70%였으며, Tm 값은 55 °C에서 68 °C 사이가 되도록 프라이머를 제작하였다. 제작된 프라이머는 각각 50 μ M로 희석한 뒤, 정방향 프라이머(forward primer)와 역방향 프라이머(reverse primer)를 혼합하였다.

2.4. Multiplex PCR

Real-time PCR을 통해 확인한 DNA의 농도를 0.05 ng/ μ L로 희석하여 multiplex PCR을 수행하였다. 프라이머는 50-plex 프라이머로, 총 농도는 50 μ M, 각 프라이머당 농도는 0.1 μ M이었다. PCR 튜브에 Phusion U multiplex PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) 25 μ L, 혼합된 50-plex 정방향 및 역방향 프라이머 각 5 μ L씩, DNA (농도 0.05 ng/ μ L) 1 μ L, 증류수 14 μ L를 넣고 총 50 μ L가 되도록 하였다. 98 °C에서 15분 initial denaturation 후, 98 °C 10초, 54 °C 30초, 72 °C 20초 cycle 과정 40회를 반복하고, 72 °C에서 5분 final extension 과정을 거쳐 PCR을 수행하였다. PCR 산물의 크기는 135 bp 이상 150 bp이며, 0.5X TBE buffer, 2% agarose 겔(gel)을 이용해

100 V에서 15분 동안 전기영동을 하여 확인하였다.

2.5. 겔 추출 방법(Gel extraction)

PCR 증폭 산물을 정제하기 위하여 겔 추출 방법을 수행하였다. 전기영동을 마친 겔의 135~150 bp의 증폭 산물 부분을 잘라 QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany)를 이용해 겔 추출을 수행하였다. 겔 추출 결과로 얻어진 최종 산물의 부피가 30 μ L가 되도록 하였고, Qubit™ Quantification Assay Kit (dsDNA, High Sensitivity)를 이용해 Qubit™ 2.0 시스템(Thermo Fisher Scientific, Waltham, MA, USA)으로 정량하였다.

2.6. NGS 분석

NGS 분석을 위한 라이브러리는 일루미나 시스템에 적합한 short read 방식으로 제작되었다. TruSeq DNA nano Kit (Illumina, San Diego, CA)의 매뉴얼대로 라이브러리 제작을 진행하였다. 제작된 라이브러리는 NovaSeq 6000 시스템(Illumina, San Diego, CA)을 이용해 NGS 분석을 진행하였다.

NGS 결과로 얻은 fastq 파일은 리눅스 기반 운영체제에서 프라이머를 설계할 때 사용했던 201 bp sequence를 참조서열(GRCh38)로 하여 mapping 되었다. Bowtie2 (2.2.5), BWA (0.1.17)를 이용하여 참조서열에 정렬하였고, Samtools (1.13)를 이용해 파일을 변환하였다. 결과는 Integrative Genomics Viewer (IGV, 2.6.1)를 이용하여 시각화하였다.

2.7. 패널 검증을 위한 추가 시료 NGS 분석

패널 검증을 위한 14개의 추가 시료의 라이브러리 제작을 위해 NEBNext Ultra II DNA library prep kit for Illumina (New England Biolabs, Ipswich, MA)의 매뉴얼에 따라 라이브러리를 제작하였다. 제작된 라이브러리는 MiSeq FGx 시스템(Illumina, San Diego, CA)을 이용해 NGS 분석을 진행하였다.

NGS 결과로 얻은 fastq 파일은 리눅스 기반 운영체제에서 프라이머를 설계할 때 사용했던 201 bp sequence를 참조서열(GRCh38)로 mapping하였다. Bowtie2 (2.2.5)를 이용하여 참조서열에 정렬하였고, 정렬된 파일은 Samtools (1.13)을 이용해 변환하였다. 각 SNP는 FreeBayes (1.0.0)을 이용해 변이가 있는 유전형을 vcf 파일로 추출하고 Integrative Genomics Viewer(IGV, 2.6.1)를 이용하여 시각화하였다.

Table 1. Selected 50 SNP markers associated with external visible characteristics(EVCs) and behavior

No.	SNP	Gene	Chr.	Position	Phenotype	(Allele) Phenotype
1	rs671	ALDH2	12	111803962	Alcohol metabolism	(A) Alcohol sensitive(blush)
2	rs629725	FST	5	53335237	Acne	(T) Acne skin
3	rs633715	SEC16B	1	177883445	Body shape	(C) Obesity
4	rs788790	LGR6	1	202320478	Acne	(A) Acne skin
5	rs922452	EDAR	2	108927427	Androgenetic alopecia	(A) Thicker hair
6	rs1126809	TYR	11	89284793	Skin Color	(A) Lighter skin
7	rs1129038	HERC2	15	28111713	Skin Color	(G) Darker skin
8	rs1160312	LINC01432	20	22069865	Androgenetic alopecia	(A) Alopecia
9	rs1229984	ADH1B	4	99318162	Alcohol metabolism	(A) Alcohol sensitive
10	rs1334576	RREB1	6	7211585	Body shape	(G) Abdominal obesity
11	rs1451240	CHRNA3	8	42691568	Nicotine metabolism	(C) Nicotine dependence
12	rs1470608	OCA2	15	28042975	Skin Color	(A) Lighter skin
13	rs1545397	OCA2	15	27942626	Skin Color	(T) Intermediate-darker skin
14	rs1667394	HERC2	15	28285036	Skin Color	(G) Darker skin
15	rs1800414	OCA2	15	27951891	Skin Color	(G) Lighter skin
16	rs2073963	HDAC9	7	18838251	Androgenetic alopecia	(G) Alopecia
17	rs2238289	HERC2	15	28208069	Skin Color	(C) Lighter skin
18	rs2241423	MAP2K5	15	67794500	Body shape	(G) Gain body weight
19	rs2273500	CHRNA4	20	63355597	Nicotine metabolism	(C) Nicotine dependence
20	rs2294239	ZNRF3	22	29053489	Body shape	(A) Abdominal obesity
21	rs2531995	ADCY9	16	3963466	Body shape	(A) Obesity
22	rs2645294	WARS2	1	119031964	Body shape	(A) Abdominal obesity
23	rs2733832	TYRP1	9	12704725	Skin Color	(T) Lighter skin
24	rs3114908	ANKRD11	16	89317317	Skin Color	(A) Intermediate skin
25	rs3212355	MC1R	16	89917970	Skin Color	(T) Darker skin
26	rs3736712	WDR27	6	169557868	Eye shape	(C) Longer eye tail
27	rs3827760	EDAR	2	108897145	Hair thickness	(C) Thicker hair
28	rs4752566	FGFR2	10	121508117	Hair thickness	(A) Thicker hair
29	rs4765219	CCDC92	12	123955563	Body shape	(C) Abdominal obesity
30	rs6059655	RALY	20	34077942	Skin Color	(A) Lighter skin
31	rs6119471	ASIP	20	34197406	Skin Color	(C) Intermediate skin
32	rs6469937	SNTB1	8	120598198	Refractive error	(T) Refractive error
33	rs6497292	HERC2	15	28251049	Skin Color	(G) Darker skin
34	rs6548238	TMEM18	2	634905	Body shape	(C) Gain body weight
35	rs7498665	SH2B1	16	28871920	Body shape	(G) Gain body weight
36	rs7567283	OSR1-WDR35	2	19532530	Face shape	(G) Small face width
37	rs7988412	GTF3A	13	27426145	Body shape	(T) Gain body weight
38	rs8051733	DEFF8	16	89957798	Skin Color	(G) Lighter skin
39	rs9939609	FTO	16	53786615	Body shape	(A) Gain body weight
40	rs10502861	SETBP1	18	45220183	Androgenetic alopecia	(C) Alopecia
41	rs10756819	BNC2	9	16858086	Skin Color	(C) Darker skin
42	rs10838738	MTCH2	11	47641497	Body shape	(C) Gain body weight
43	rs10938397	GNPDA2	4	45180510	Body shape	(G) Gain body weight
44	rs11030104	BDNF	11	27662970	Body shape	(T) Gain body weight
45	rs11084753	KCTD15	19	33831232	Body shape	(G) Gain body weight
46	rs12441727	OCA2	15	28026629	Skin Color	(T) Intermediate skin
47	rs12565727	TARDBP	1	10973025	Androgenetic alopecia	(A) Alopecia
48	rs13382811	ZEB2	2	144466053	Refractive error	(A) Refractive error
49	rs17128291	SLC24A4	14	92416482	Skin Color	(G) Lighter skin
50	rs56113850	CYP24A6	19	40847202	Nicotine metabolism	(C) Nicotine dependence

3. 결과 및 고찰

3.1. SNP 마커의 선별

한국인을 비롯한 동아시아인에게서 효과적인 표현형 관련 SNP 마커를 선별하기 위해 외형 추정 특성 및 행동 습관 관련 SNP를 탐색하였다. 눈동자 색, 머리카락 색, 피부 색 관련 SNP 43개,¹¹⁻¹³ 얼굴 모양 관련 SNP 1개,¹⁴ 눈 모양 관련 SNP 1개,¹⁴ 코 모양 관련 SNP 11개,¹⁵⁻¹⁷ 머리카락 굵기 관련 SNP 3개,^{18,19} 탈모 관련 SNP 14개,²⁰⁻²⁴ 비만 관련 SNP 25개,²⁵⁻²⁷ 고도근시 관련 SNP 2개,^{28,29} 여드름 발생 관련 SNP 3개,³⁰ 알코올 대사 관련 SNP 2개,³¹⁻³³ 니코틴 대사 관련 SNP 3개³⁴⁻³⁶를 탐색했다. 1차 선별한 SNP의 수는 총 108개였다. 이 중 한국인에게서 발생 빈도가 낮아 Minor 대립유전자 빈도가 0.01 미만 이거나, 표현형 추정이 어려운 경우, 그리고 이미 널리 사용 중인 눈동자 색, 머리카락 색 관련 SNP는 제외하였다. 실제 범죄 수사 및 신원확인에 정보를 제공할 수 있을 것으로 판단되는 총 50개의 외형 추정 및 행동 습관 관련 SNP 마커를 선별하였다. 최종 선별된 SNP 마커는 피부 색 관련 17개, 탈모 및 머리카락 굵기 관련 7개, 고도근시 관련 2개, 얼굴 모양 관련 1개, 눈 모양 관련 1개, 체형 관련 15개, 여드름 관련 2개, 알코올 대사 관련 2개, 니코틴 의존 관련 3개이다(Table 1).

3.2. Targeted amplicon NGS를 통한 50개의 SNP 부위 확인

선별된 SNP를 분석하기 위해서 targeted amplicon NGS 방법을 이용하였다. Targeted amplicon NGS는 전체 유전체의 염기서열을 분석하는 것이 아니라, 확인하고자 하는 부위를 중점적으로 분석하는 방법이다. 먼저, 분석하고자 하는 특정 DNA 영역을 증폭하고, 증폭된 amplicon에 어댑터를 부착하여 시퀀싱(sequencing)하는 방법으로, 특정 영역의 유전적 변이를 효율적으로 분석할 수 있다. 이러한 targeted amplicon NGS의 경우, 특정 부위만 시퀀싱 하므로 전체 유전체 시퀀싱에 비해 비용을 절감할 수 있으며, 특정 유전자 영역만 분석하므로 더 높은 coverage 데이터를 얻을 수 있는 장점이 있다.³⁷

PCR amplicon을 이용한 targeted amplicon NGS 방식은 multiplex PCR로 타겟(target) 서열을 증폭하기 때문에, 프라이머 수가 증가할수록 프라이머 dimer의 수도 증가하여 타겟 서열 수를 확장하는 데 어려움이 있다. Hybridization-capture 방식을 사용했을 때 타겟

부위를 수백 개 이상으로 확장할 수 있지만, amplicon 방식에 비해 라이브러리 제작에 많은 시간이 소요되고 높은 농도의 DNA가 필요하다.³⁸ 대부분 현장 증거물에서 유래된 DNA 시료는 농도가 낮기 때문에, 50개의 SNP 부위를 선택하여 PCR amplicon 방식으로 라이브러리를 제작하여 targeted amplicon NGS를 시행하고자 하였다. 따라서 50개 SNP 부위에 대한 multiplex PCR용 프라이머를 제작하였다(Table 2). 실제 제작된 프라이머를 이용하여 multiplex PCR을 수행한 다음 NGS 방법을 이용하여 SNP 서열을 분석하였다.

NGS 분석 결과 총 15,618,078의 read가 생성되었으며, 이들 중 14,397,943 (92.19%)의 read가 mapping 되었다. 그리고 50개의 SNP 마커에 대해 모두 read가 확인되었으므로, 50개 SNP 부위가 multiplex PCR로 증폭된 것을 확인할 수 있었다(Fig. 1, Table 3). 하지만 비만 관련 SNP인 rs633715 마커의 경우 100 read 미만으로 확인되었으므로, 감도가 더 높은 NGS 장비/시약의 사용 또는 프라이머의 교체 등의 추가적인 보완 있어야 할 것으로 판단된다.

50개의 SNP에 대한 mapping을 통하여 각 SNP 변이의 read수를 확인하였고, 이를 통해 각 SNP 유전형에 대한 빈도를 확인할 수 있었다. 본 연구에서 확인한 SNP들은 germline 변이로, 부계와 모계에서 하나씩 유전되므로 동형접합 또는 이형접합의 형태를 가진다. 일반적으로 각 대립유전자 빈도는 1일 때 동형접합으로, 0.5일 때 이형접합으로 판단한다. 하지만 PCR 과정에서 두 대립유전자의 불균형한 증폭 및 시퀀싱 오류로 인한 편향 때문에, 더 넓은 범위인 0.2~0.8 대립유전자 빈도에서 이형접합으로 판단될 수 있다.³⁹

3.3. NGS 패널 검증을 위한 추가 시료 NGS 분석

본 연구에서 개발된 NGS 패널을 이용하여 14개의 추가 시료를 분석하였다. 14개 시료에서 multiplex PCR을 통해 50개의 SNP 부위를 증폭하여 라이브러리를 제작하였다. NGS 분석결과 일루미나 시스템 플랫폼의 종류와 관계없이 모든 SNP에서 read가 확인되었고 50개 SNP의 유전형과 대립유전자 빈도를 확인할 수 있었다(Table 4). 이후에는 표현형을 알고 있는 200명의 시료를 대상으로 SNP 유전형을 확인하고, 표현형과의 연관 정도 분석 등을 통하여 외형 및 행동 습관 추정 과정의 확립을 이룰 수 있을 것으로 기대된다.

Table 2. Primer sequences to amplify selected 50 SNP markers

SNP	Primer_Forward	Primer_Reverse
rs671	GGAGCCCAGTCACCCTTTG	AGACCCTCAAGCCCCAACAG
rs629725	TCTTGTGGAGTCTGTTGATGCC	GTTGGCAGCAGTGACAATAAT
rs633715	TAGCATTGAAAAGCATAACAGGCA	TTCAAGGTCTGGTACCTGCTTA
rs788790	GATAGGCCGCTGACCCTTC	CAGCTCTTACCGCCTCACTC
rs922452	GCAGCTGCCCCACTAATTC	GACCACGTGCTTGTTCTGATA
rs1126809	GTCTGAATAACCTTTTCTCTGCG	TGTACAGTGGTATAAAAAGGAACCA
rs1129038	CGCTTCTCATCAGACACACCA	ACTCCTTTGCTTCGGACTCTAC
rs1160312	ACATAACCATCTCCCGGGGC	TCAGGCCGGAGAATCACACTC
rs1229984	TCTCCAACACTCTCCACGATG	TTCTGAATCTGAACAGCTTCTCTT
rs1334576	ATGTGGGAGACCTTCATGACTT	GTCTCCAGTCCATACTTGCAAAC
rs1451240	GACACTTCCTGGTCCCTCCGAC	TCGCCTCTCCTTTCATCCTTCA
rs1470608	GCCCCCTAATTGGAAAATATGTTAGG	ACAAATTGTCTGGAAGCATTTTAAAAACAG
rs1545397	TCTGGAATTGGATACTGACAATGGTT	AAATCATGGGGGAGAGAGAATGAC
rs1667394	ATCCACCATTAAGACGCAGCAATTC	TATCAGCTGTAGAGAGAGACTTTGAGG
rs1800414	CTGTCAGGCATTTGGCGAG	AAATCTGTCGTGATTCCAGTTGC
rs2073963	GCACGTGTGGTAAGCACTTGG	TCCTGAAACCATTGAGGACAAG
rs2238289	CGGCCCAAAGCACAGACTT	GGGAACATGAAGATTTCCAGTTG
rs2241423	AGGAGGGGTCTGTTTCAGATGG	GGGAAAATTTGGAAGAAGAGCG
rs2273500	ATGGGGTCTGATGGCGAAAA	GGCTATGGCACCTGGACAC
rs2294239	TCTCAGCACAAATGGAGCCC	CAATCATCAGGGGAGGGAGC
rs2531995	TCTGTGGTTCTGCACTGAGG	GCAGACTCACCCAGAGCTTA
rs2645294	ATGGGGAAATGTGACCACGG	TGCTGAAAGGCAGCGTTTTT
rs2733832	TTTGATGAATGGCTGAGGAGATACA	TTTTGAAAGTGTTTTTATGCGCTGAG
rs3114908	GCCTCCCATAAAGGGGTAC	GATCCAGGTGCTGCTGACTG
rs3212355	AGTGAACCCAGGAAGATGCC	ACAGCCACATCAAAGGCAGA
rs3736712	TTAAAACTTGCATTTCCCTCCACTTAATA	TCCCATTTTTATATTCAAGTCAATATGTTTTTCA
rs3827760	TCAAAGAGTTGCATGCCGTC	GCCTCGAGAAGACTAGCCG
rs4752566	AGGATGTACACAGTTCATATGCCAAATAC	CCTGGTTGCACACAGTTTGA
rs4765219	CTAGGCCAAGGAGATGTGTTT	CTGTACTATAATTAGGCAGACCCA
rs6059655	CCCACATTTTACCCTGTGAGGA	TGCTGAAAAGGAGAAAAGCTGC
rs6119471	AGCAGAGATCGTGCCACTG	GGAGGCTAACCCGAAGGAAG
rs6469937	AGTCCCCTTCTGGCTCTCA	CCAGATAAGCCCTGAGTTCC
rs6497292	ACGTACATCCTTATTACAAATCTGC	GGTGGCTCTGTGCTGTCC
rs6548238	TCTTGTACAGGCGACAGACC	GCCAGTTGTAGGGATGAACGA
rs7498665	GAACGTCTCCCCCAGAGTTG	GCACACTCACACCTCCGGTA
rs7567283	ACATCCCCATTTTACAGAATGAAC	CATGAGCATAATAAAAACAGTTGCTG
rs7988412	TTTATGCTTTGCCCTCTGC	TGGTAGCGCCAGTATCCAAA
rs8051733	CAGAACTGTGGAAGTAGGAGGAC	AATAGGGAAACCTCCCGCA
rs939609	TTATTATTAGGTTTCTTGGCACTGC	AGCTATTTGCATTTAGTTTGTATG
rs10502861	GCTCTATCAGCTGAAGTCATTCC	GCTAAGACCAGGTGATTTGGTT
rs10756819	AACTGTAAAGCAAGCTCATGTTTC	TGACTAGAAAAACACCAACAGAGA
rs10838738	AGAAAAGTAGACGGCGAGACA	TCCAGTTCCTTTGCTATATCCCTG
rs10938397	TGTACACACACCAAAATGTTTTACTTTAC	AATTTGCCAAAGGACATAGCTGAATG
rs11030104	AGCCATGACCAACTTCTTGAG	TCCCAAGTCTCTTATTGATCTGTTA
rs11084753	TCAGCTATGAAAAGAGATGAAGCG	AACCTTAGGTGATCCACCCG
rs12441727	AGTGGGAAGAGACAGCTCCA	AATCTGGGAGGTACACGGA
rs12565727	TCTGTTTACCTGGGGCATTGG	CACTGCACCTGGCCTGACTTTA
rs13382811	CTCCAGGGCAGAACTTTCCA	CCTTGTGAAGATGACAAAAGGCA
rs17128291	AGGAACTGAGGCCAGCAC	TTGGACCCATCACCTCTCTTC
rs56113850	GCATCTGTCTCGTTGTTTAGG	CTCGTTTAAATACCTGAAACCTGG

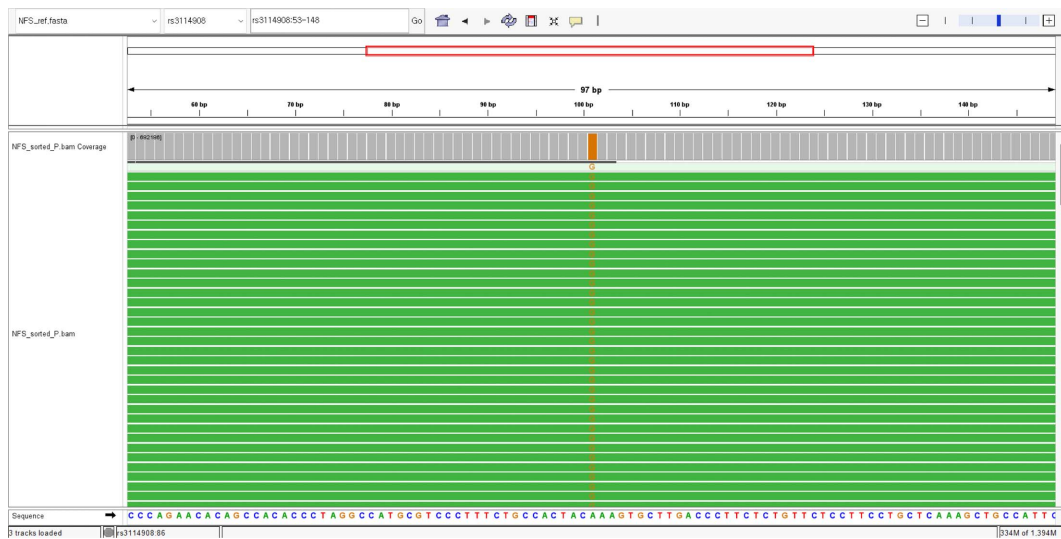


Fig. 1. Visualization of read sequence of rs3114908 using IGV. Target SNP is located at 101 bp.

Table 3. Targeted NGS results for 50 SNP markers

Phenotype	rs#	Genotype 1	Frequency 1	Genotype 2	Frequency 2	Total Count
Abdominal obesity	rs2294239	A	0.50	G	0.50	745184
Abdominal obesity	rs4765219	C	0.55	A	0.45	151437
Abdominal obesity	rs2645294	A	0.83	G	0.17	4166
Abdominal obesity	rs1334576	A	1.00	-		106710
Alcohol Use Disorder	rs1229984	A	0.79	G	0.21	88296
Alcohol Use Disorder	rs671	A	0.62	G	0.38	543160
Body Mass Index	rs7988412	C	1.00	-		26348
Body Mass Index	rs11030104	T	0.60	C	0.40	39743
Body Mass Index	rs2241423	A	1.00	-		24895
Body Mass Index	rs9939609	T	0.89	A	0.11	642040
Body Mass Index	rs6548238	C	1.00	-		388307
Body Mass Index	rs10938397	A	1.00	-		418
Body Mass Index	rs7498665	A	1.00	-		93905
Body Mass Index	rs10838738	T	0.60	C	0.39	5225
Body Mass Index	rs11084753	G	0.60	A	0.40	9926
Eye shape	rs3736712	T	1.00	-		12511
Face shape	rs7567283	A	0.99	G	0.01	2211
Hair thickness	rs3827760	C	1.00	-		31998
Hair thickness	rs922452	A	1.00	-		251147
Hair thickness	rs4752566	A	0.90	C	0.09	47538
Male Pattern Alopecia	rs1160312	G	1.00	-		219251
Male Pattern Alopecia	rs2073963	T	0.54	G	0.46	5673
Male Pattern Alopecia	rs10502861	C	1.00	-		1293958
Male Pattern Alopecia	rs12565727	A	0.59	G	0.41	1166148
Myopia	rs6469937	T	1.00	-		14204
Myopia	rs13382811	G	0.65	A	0.35	271311
Nicotine dependence	rs1451240	T	0.58	C	0.41	737
Nicotine dependence	rs2273500	T	1.00	-		117359

Table 3. Continued

Phenotype	rs#	Genotype 1	Frequency 1	Genotype 2	Frequency 2	Total Count
Nicotine dependence	rs56113850	C	0.55	T	0.45	57959
Obesity	rs633715	T	1.00	-		32
Obesity	rs2531995	A	0.51	G	0.48	499246
Severe acne	rs629725	C	0.79	T	0.21	415249
Severe acne	rs788790	A	0.97	C	0.03	1884864
Skin	rs2733832	C	1.00	-		1757
Skin	rs3114908	G	1.00	-		682183
Skin	rs10756819	T	0.98	C	0.02	773
Skin	rs1126809	G	1.00	-		58037
Skin	rs17128291	A	1.00	-		153614
Skin	rs1667394	G	0.96	A	0.04	56393
Skin	rs2238289	T	0.70	C	0.30	438664
Skin	rs1129038	G	1.00	-		106486
Skin	rs6497292	G	0.62	A	0.38	179251
Skin	rs1800414	G	0.97	A	0.03	865302
Skin	rs12441727	C	1.00	-		164754
Skin	rs1470608	A	0.76	C	0.24	1374
Skin	rs1545397	A	1.00	-		398
Skin	rs8051733	G	0.73	A	0.27	43787
Skin	rs3212355	C	1.00	-		1124829
Skin	rs6119471	C	1.00	-		20440
Skin	rs6059655	G	1.00	-		456086

Table 4. Genotypes obtained with targeted NGS results from 14 additional samples

SNP	Sample_Genotype													
	NFS_1	NFS_2	NFS_3	NFS_4	NFS_5	NFS_6	NFS_7	NFS_8	NFS_9	NFS_10	NFS_11	NFS_12	NFS_13	NFS_14
rs1160312	AA	AG	GG	GG	GG	GG	AG	AA	AA	GG	AG	AG	GG	GG
rs3827760	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	CC	CC	CC	CC
rs922452	AA	AA	AA	AA	AA	AA	AA	AA	AA	GA	AA	AA	AA	AA
rs4752566	AA	CA	CA	CA	AA	AA	CC	CA	CA	AA	CA	AA	AA	AA
rs7988412	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
rs11030104	TC	CC	CC	TC	CC	TC	TC	CC	TC	TC	TT	TC	TT	TC
rs2241423	AA	GA	AA	GA	AA	AA	AA	GA	GA	GG	GA	GA	AA	AA
rs6469937	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	CT
rs13382811	GG	GA	AA	GA	GG	GA	GG	GG	GG	GA	GA	GG	GA	GG
rs7567283	AA	GA	GA	GA	GA	GA	GG	AA	GA	AA	GA	AA	AA	GA
rs3736712	TT	TT	TT	TT	TT	TT	CT	CT	TT	CT	TT	TT	TT	CT
rs1229984	AA	AA	AA	AA	AG	AA	AA	AG	AG	AA	AA	AG	AG	AG
rs671	GG	AA	AA	AA	GG	GA	GG	GG	AA	GG	GG	GG	GA	GA
rs629725	CC	CC	CC	CC	CC	CC	TC	CC	CC	CC	CC	TC	TC	TC
rs788790	AA	CA	AA	CA	AA	CA	AA	CA	CA	CA	AA	CA	AA	CA
rs2073963	TG	TT	TT	TT	TT	TT	TG	TG	TG	TG	TT	TG	TG	TG
rs10502861	CC	CT	CC	CC	CC	CC	CC	CC	CC	CC	CT	CC	CC	CC
rs12565727	AG	AG	AG	AA	AA	AG	AA	AA	AA	AG	AA	AA	AG	AG
rs633715	TC	CC	TT	CC	TT	CC	TT	TT	TC	TT	TC	TT	TT	TT
rs2531995	GG	GA	GA	AA	GG	GA	GG	GA	GG	GA	AA	GG	GA	GA

Table 3. Continued

SNP	Sample_Genotype													
	NFS_1	NFS_2	NFS_3	NFS_4	NFS_5	NFS_6	NFS_7	NFS_8	NFS_9	NFS_10	NFS_11	NFS_12	NFS_13	NFS_14
rs9939609	TT	TT	TT	TT	TT	TT	TT	TT	TT	TA	TT	TT	TA	TT
rs6548238	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
rs10938397	AG	AA	AA	AA	AA	AA	AA	AG	AA	AA	AA	AA	AA	AG
rs7498665	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs10838738	TC	TC	TC	CC	TC	TC	CC	TT	TC	TC	TC	CC	TC	TC
rs11084753	AA	AG	AG	AG	AG	AA	AA	AG	GG	AG	AG	AA	AA	AA
rs2294239	AG	AG	AG	GG	AA	AG	AA	AA	AG	AG	AG	AG	AG	AG
rs4765219	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CA	CC	CC	CC
rs2645294	AA	GA	GG	GA	AA	GA	AA	AA	AA	GA	AA	GA	AA	AA
rs1334576	GG	GG	GA	GA	GA	GA	GG	GA	AA	AA	GG	AA	AA	GG
rs2733832	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
rs1451240	CC	CC	TC	TC	CC	CC	CC	CC	TC	TC	CC	TC	TC	TC
rs2273500	TC	TC	TC	TT	TT	TT	TT	TT	TT	TC	TT	TC	TT	TC
rs56113850	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC	TC
rs3114908	GG	AA	AG	AA	AG	AG	GG	AG	GG	AG	AG	GG	GG	AA
rs10756819	TT	TT	TT	TT	CT	TT	CT	TT	TT	TT	TT	CT	TT	CT
rs1126809	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
rs17128291	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs1667394	GG	GA	GA	GG	GG	GG	GG	GG	GG	GA	GG	GG	GG	GA
rs2238289	TT	TC	TT	TC	TT	TC	TT	TT	TT	TT	TT	TT	TC	TT
rs1129038	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
rs6497292	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG	AG	GG	AG	AA
rs1800414	AG	GG	AG	GG	AG	AG	AG	AA	GG	AG	GG	AG	GG	GG
rs12441727	CT	CC	CT	CC	CT	CT	CT	CT	CC	CT	CT	CC	CC	CC
rs1470608	AA	CA	CA	AA	AA	AA	CA	CA	AA	CA	AA	AA	AA	AA
rs1545397	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs8051733	AA	AA	AG	AA	AA	AG	AA	AA	AG	AA	AA	AG	AG	AG
rs3212355	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
rs6119471	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
rs6059655	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG

4. 결 론

본 연구에서는 법과학 시료의 분석을 위한 50개의 외형 추정 및 행동 습관 관련 SNP를 선별하였고, targeted amplicon NGS 방법을 이용하여 각 SNP의 유전형을 확인할 수 있는 NGS 패널을 개발하였다. 본 실험 기법은 DNA의 양이 적거나 고도로 분해되어 STR 분석이 어려운 시료에서 효과적으로 분석을 수행할 수 있어 개인을 추정하는 데 법과학적인 증거를 제공할 수 있을 것으로 사료된다. 향후, 개발된 NGS 패널을 이용하여 더 많은 시료를 분석하고, 머신러닝을 통해 대립유전자형 빈도에 따라 해당 유전자형과 표현형의 상관관계를 분석하여 표현형 예측 성능을

평가할 예정이다. 이러한 데이터를 기반으로 배경 정보가 없는 시료에서 유전자형을 통한 실제 표현형 추정에 적용할 예정이다. 이러한 실험 기법은 다양한 SNP 마커를 활용한 분석에 적용할 수 있을 것으로 생각되며, 범죄 현장 증거물에서의 용의자 추정 및 신원 불상 변사자의 신원 추정에서 중요한 증거를 제공할 것으로 기대된다.

감사의 글

이 논문은 행정안전부 주관 국립과학수사연구원 중장기과학수사감정기법연구개발(R&D)사업의 지원을 받아 수행한 연구임(NFS2024DNA02).

References

1. J. M. Butler, M. D. Coble, and P. M. Vallone, *Forensic Sci. Med. Pathol.*, **3**, 200-205 (2007). <https://doi.org/10.1007/s12024-007-0018-1>
2. R. Thompson, S. Zoppis, and B. McCord, *Methods Mol. Biol.*, **830**, 3-16 (2012). https://doi.org/10.1007/978-1-61779-461-2_1
3. M. Kayser and P. M. Schneider, *Forensic Sci. Int. Genet.*, **3**(3), 154-161 (2009). <https://doi.org/10.1016/j.fsigen.2009.01.012>
4. P. M. Schneider, B. Prainsack, and M. Kayser, *Dtsch Arztebl Int.*, **51-52**(51-52), 873-880 (2019). <https://doi.org/10.3238/arztebl.2019.0873>
5. M. Kayser, *Forensic Sci. Int. Genet.*, **18**, 33-48 (2015). <https://doi.org/10.1016/j.fsigen.2015.02.003>
6. M. Kayser and P. Knijff, *Nat. Rev. Genet.*, **12**, 179-192 (2011). <https://doi.org/10.1038/nrg2952>
7. B. Sobrino and A. Carracedo, *Methods Mol. Biol.*, **297**, 107-126 (2005). <https://doi.org/10.1385/1-59259-867-6:107>
8. M. Kowalczyk, E. Zawadzka, D. Szewczuk, M. Gryzinska, and A. Jakubczak, *Med. Sci. Law*, **58**(4), 201-209 (2018). <https://doi.org/10.1177/0025802418803852>
9. H. J. Seo, S. H. Cho, J. H. Lee, S. H. Lyoo, M. Y. Kim, and S. D. Lee, *Korean J. Leg. Med.*, **14**, 23-31 (2017). <https://doi.org/10.7580/kjlm.2017.41.2.23>
10. P. M. Schneider, B. Prainsack, and M. Kayser, *Dtsch Arztebl Int.*, **116**(51-52), 873-880 (2019). <https://doi.org/10.3238/arztebl.2019.0873>
11. R. A. Strum, *Hum. Mol. Genet.*, **18**(R1), R9-R17 (2009). <https://doi.org/10.1093/hmg/ddp003>
12. S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, and M. Kayser, *Forensic Sci. Int. Genet.*, **7**(1), 98-115 (2013). <https://doi.org/10.1016/j.fsigen.2012.07.005>
13. L. Chaitanya, et al., *Forensic Sci. Int. Genet.*, **35**, 123-135 (2018). <https://doi.org/10.1016/j.fsigen.2018.04.004>
14. S. Cha, et al., *BMC Genomics*, **19**, 481 (2018). <https://doi.org/10.1186/s12864-018-4865-9>
15. F. Liu, et al., *PLoS Genet.*, **8**, e1002932 (2012). <https://doi.org/10.1371/journal.pgen.1002932>
16. L. Paternoster, et al., *Am. J. Hum. Genet.*, **90**(3), 478-485 (2012). <https://doi.org/10.1016/j.ajhg.2011.12.021>
17. M. K. Lee, J. R. Shaffer, E. J. Leslie, E. Orlova, J. Carlson, E. Feingold, M. L. Marazita, and S. M. Weiner, *PLoS One*, **12**, e0176566 (2017). <https://doi.org/10.1371/journal.pone.0176566>
18. A. Fujimoto et al., *Hum. Mol. Genet.*, **17**(6), 835-843 (2008). <https://doi.org/10.1093/hmg/ddm355>
19. A. Fujimoto et al., *J. Hum. Genet.*, **54**, 461-465 (2009). <https://doi.org/10.1038/jhg.2009.61>
20. J. B. Richards et al., *Nat. Genet.*, **40**(11), 1282-1284 (2008). <https://doi.org/10.1038/ng.255>
21. L. Petukhova et al., *Nature*, **466**(7302), 113-117 (2010). <https://doi.org/10.1038/nature09114>
22. K. K. G. John, et al., *J. Invest. Dermatol.*, **131**(5), 1169-1172 (2011). <https://doi.org/10.1038/jid.2010.427>
23. R. Li, et al., *PLoS Genet.*, **8**, e1002746 (2012). <https://doi.org/10.1371/journal.pgen.1002746>
24. S. P. Hagenaars, W. D. Hill, S. E. Harris, S. J. Ritchie, G. Davies, D. C. Liewald, C. R. Gale, D. J. Porteous, I. J. Deary, and R. E. Marioni, *PLoS Genet.*, **13**(2), e1006594 (2017). <https://doi.org/10.1371/journal.pgen.1006594>
25. C. J. Willer et al., *Nat. Genet.*, **41**(1), 25-34 (2009). <https://doi.org/10.1038/ng.287>
26. S. I. Berndt et al., *Nat. Genet.*, **45**, 501-512 (2013). <https://doi.org/10.1038/ng.2606>
27. D. Shungin et al., *Nature*, **518**, 187-196 (2015). <https://doi.org/10.1038/nature14132>
28. C. C. Khor et al., *Hum. Mol. Genet.*, **22**(25), 5288-5294 (2013). <https://doi.org/10.1093/hmg/ddt385>
29. V. J. M. Verhoeven et al., *Nat. Genet.*, **45**(3), 314-318 (2013). <https://doi.org/10.1038/ng.2554>
30. C. Petridis et al., *Nat. Commun.*, **9**, 5075 (2018). <https://doi.org/10.1038/s41467-018-07459-5>
31. T. L. Wall, S. M. Horn, M. L. Johnson, T. L. Smith, and L. G. Carr, *J. Stud. Alcohol*, **61**(1), 13-17 (2000). <https://doi.org/10.15288/jsa.2000.61.13>
32. R. Sherva, J. P. Rice, R. J. Neuman, N. Rochberg N. L. Saccone, and L. Bierut, *Alcohol Clin. Exp. Res.*, **33**(5), 848-857 (2009). <https://doi.org/10.1111/j.1530-0277.2009.00904.x>
33. J. Gelernter, et al., *Mol. Psychiatry*, **19**, 41-49 (2014). <https://doi.org/10.1038/mp.2013.145>
34. B. Hancock, et al., *Transl. Psychiatry*, **5**, e651 (2015). <https://doi.org/10.1038/tp.2015.149>
35. A. Loukola, et al., *PLoS Genet.*, **11**, e1005498 (2015).

- <https://doi.org/10.1371/journal.pgen.1005498>.eCollection 2015
36. Y. M. Patel, D. O. Stram, L. R. Wilkens, S. S. L. Park, B. E. Henderson, L. L. Marchand, C. A. Haiman, and S. E. Murphy, *Cancer Epidemiology Biomarkers Prev.*, **24**(1), 119-127 (2015). <https://doi.org/10.1158/1055-9965.EPI-14-0815>
37. A. C. Jäger et al., *Forensic Sci. Int. Genet.*, **28**, 52-70 (2017). <https://doi.org/10.1016/j.fsigen.2017.01.011>
38. N. G. Xie, M. X. Wang, P. Song, Y. Wang, Y. Yang, J. Luo, S. Ren, and D. Y. Zhang, *Nat. Commun.*, **13**, 1880 (2022). <https://doi.org/10.1038/s41467-022-29500-4>
39. R. Nielsen, J. S. Paul, A. Albrechtsen, and Y. S. Song, *Nat. Rev. Genet.*, **12**, 443-451 (2011). <https://doi.org/10.1038/nrg2986>

Authors' Positions

Hee-Yeon Park : Scientific Officer
Yoonji Noh : Researcher
Eung-Soo Kim : Senior Scientific Officer
Hyun-Chul Park : Senior Scientific Officer