

**Application of Identity and Ancestry Single Nucleotide
Polymorphisms to Forensic Casework Analysis in Qatar**

by

Waad Rashed Sh A Al-Dosari

A thesis submitted in partial fulfilment for the requirements for the degree of
Doctor of Philosophy at the University of Central Lancashire

September 2020



STUDENT DECLARATION FORM

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution. I declare that no material contained in the thesis, except where otherwise stated, has been used in any other submission for an academic award and is solely my own work.

Waad Rashed Sh A Al-Dosari

Type of Award: Doctor of Philosophy

School: School of Forensic and Applied Sciences

ABSTRACT

Within forensic genetics, the typing single nucleotide polymorphisms (SNPs) is an alternative to short tandem repeats (STRs). It has potential for analysing forensic samples that show high levels of degradation and can also be used to obtain information on aspects relating to phenotype and ancestry. The availability of massively parallel sequencing (MPS) platforms and commercial ready SNP panels makes the implementation of SNP testing feasible for a large number of laboratories.

This study was design to evaluate the massive parallel sequencing workflow within the Forensic Laboratory of Qatar using two ready-to-use MPS panels, namely, the Precision ID Ancestry Panel and the Precision ID Identity Panel with the Ion Chef system for template and library preparation and the Ion Torrent PGM sequencer. Both panels were evaluated using population samples and forensic samples.

The Precision ID Ancestry panel comprises 165 Autosomal SNPs that can be used to infer the biogeographical origin of unknown samples. A total of 300 unrelated Qatari individuals collected from the eight official municipalities were analysed for this study. All loci proved to be in HWE after applying Bonferroni correction ($p < 0.0003$). The forensic parameters were estimated to be combined probability of match (CPM) and the combined power of exclusion (CPE) was 1.35×10^{-42} and 0.999998, respectively.

From the original 300 samples, 105 samples were profiled using Precision ID Identity Panel that includes 90 autosomal SNPs and 34 upper Y-Clade SNPs. The CMP and the CPE for the 90 autosomal SNPs were 7.5674×10^{-37} and 0.999998714. There was no significant deviation from HWE across all 90 autosomal SNPs after applying Bonferroni correction ($p < 0.0006$). When the population samples were split into three geographical regions the F_{ST} values between the groups, as expected, was very low with no detectable genetic differentiation. For 34 Y-SNPs, 10 different Y-haplo-groups were observed and the most common haplogroup in Qatar population is J followed by R1a1, R2 and E. The sensitivity was assessed for both panels and full profiles could be obtained down to 0.25 ng with 21 cycles.

For the casework samples experiments, 148 real casework samples from 76 cases were collected from different security departments between 2005 and 2018. The samples varied from routine stains with high amounts of DNA to challenging samples. The

Precision ID Ancestry Panel was used to analyse 143 casework samples from the 76 cases with two workflows: manual and automated library preparation, using the Ion Chef. There were no significant differences seen between the two workflows in terms of success rates. However, the automated workflow saved considerable hands-on laboratory time.

Sixty casework samples were chosen from 46 cases (from the 76 cases above), of which 55 samples were analysed using the Precision ID Identity panel. Adding the maximum volume of template DNA and increasing the PCR cycles helped to generate useful profiles in Ancestry and Identity experiments with DNA input down to 0.12 ng and 0.06 ng, respectively. However, neither panel was as robust as the STR typing that was also tested. This finding suggested that sensitivity was reduced in comparison to the STR technology either through the chemistry of the PCR, PCR inhibition or the reduced volume of template DNA (a maximum of 6 µl template could be added to the SNP library construction whereas 15 µl could be added to the STR amplifications).

The ancestry prediction was undertaken using HID SNP Genotyper v5.2.2 as admixture prediction and population likelihood. The results obtained from the population experiment showed that most of Qatari population were, as expected, predicted to have Southwest Asian and South Asian ancestry, with a few samples predicted to have African, East Asian and European ancestry. The results supported the historical migration and settlement theory in the country. With the casework samples a different pattern of populations was seen, with increased proportion samples predicted to be of South Asian origin, which is not surprising given the demographic makeup of Qatar, with many foreigners residing in the country.

It is concluded that the MPS panels were successful in terms of DNA profiles being generated. However, they were not as robust as the existing STR systems currently employed. Additional information, in terms of ancestry, could be derived with the ancestry panel but given the composition of the Qatari population the relevance in most casework is questionable. Also, increasing the number of reference databases might improve the precision of the prediction. The additional information generated by the both panels would provide useful information in complex kinship cases.

TABLE OF CONTENTS

STUDENT DECLARATION FORM	i
ABSTRACT	ii
TABLE OF CONTENTS	iv
LIST OF ABBREVIATIONS	x
LIST OF FIGURES	xii
LIST OF TABLES	xvii
ACKNOWLEDGEMENT	xxii
CHAPTER: 1 Introduction	1
1.1 A historical perspective of forensic science and its application	1
1.2 Recent DNA typing in forensic analysis.....	2
1.3 Single nucleotide polymorphisms (SNPs).....	3
1.3.1 SNP Resources and Databases	5
1.4 Typing and detection methods	9
1.4.1 SNP typing methods	9
1.4.2 Detection methods	11
1.5 SNPs in forensic analysis	12
1.5.1 Identity SNPs	13
1.5.2 Ancestry SNPs.....	15
1.5.3 Lineage SNPs	18
1.5.4 Phenotype SNPs	19
1.6 SNP Typing Technologies	20
1.6.1 First generation Sequencing	20
1.6.2 Next generation sequencing (NGS).....	20
1.6.3 The Ion torrent PGM system.....	23
1.7 Project background	31
1.7.1 Geography and population distribution of State of Qatar	31
1.7.2 Forensic casework in Qatar	38
1.8 Project Aims	40
CHAPTER: 2 Materials and Methods	42
2.1 Overview	42
2.2 Ethics approval and permission	43
2.3 Sample collection	43
2.3.1 Qatari population samples.....	43

2.3.2 Forensic casework samples.....	45
2.4 Extraction	47
2.4.1 Extraction of population samples	48
2.4.2 Extraction of casework samples.....	48
2.5 DNA Quantitation.....	53
2.6 Library preparation and Quantitation.....	54
2.6.1 Manual library preparation and quantitation.....	54
2.7 Automation Library Preparation	60
2.7.1 Sample preparation.....	60
2.7.2 Creation of sample set	61
2.7.3 Running the Ion Chef.....	61
2.7.4 Unloading the Ion Chef and libraries dilution	62
2.8 Template preparation on the Ion Chef	63
2.8.1 Creation a planned run	63
2.8.2 Running the Ion Chef.....	64
2.9 Cleaning the Ion PGM sequencer.....	65
2.9.1 Condition the Wash 2 Bottle for First Use	66
2.9.2 Chlorite Cleaning	66
2.9.3 18 MΩ water wash	67
2.10 Initialization of the Ion PGM™ System.....	67
2.10.1 Preparation of the Wash 2 Bottle	67
2.10.2 Preparation of the Wash 1 and Wash 3 Bottles.....	67
2.10.3 Starting the Initialization of Ion PGM™	68
2.10.4 Preparation the 50 ml Reagent Bottles with dNTPs	68
2.11 Data analysis	68
2.12 Statistical analysis	69
2.12.1 SRTUCRURE Analysis	69
2.12.2 Arlequin Analysis.....	69
2.12.3 PowerStats Analysis	69
CHAPTER: 3 Evaluation of Precision ID Identity Panel in Qatar population by massively parallel sequencing using the Ion Torrent Personal Genome Machine (PGM).....	70
3.1 Introduction	70
Objectives	71
3.2 Materials and Methods.....	72
3.3 Results	72

3.3.1 Sensitivity study	72
3.3.2 Quality of the sequencing runs	75
3.3.3 Sequence Coverage	78
3.3.4 Performance of the Precision ID Identity Panel.....	79
3.3.5 Heterozygote balance and noise level	81
3.3.6 Statistical value of 90 autosomal SNPs of HID Identity SNPs in Qatar Population	83
3.3.7 Y-haplogroups	93
3.4 Discussion.....	96
3.5 Conclusion	99
CHAPTER: 4 Massively parallel sequencing of forensic samples using the Precision ID Identity panel on the Ion Torrent PGM.....	100
4.1 Introduction	100
Objectives	100
4.2 Materials and Methods.....	101
4.3 Results	101
4.3.1 Quality of the sequencing runs	102
4.3.2 Sequence Coverage.....	104
4.3.3 Performance of the Precision ID Identity Panel.....	108
4.3.4 Y-haplogroups	113
4.4 Discussion.....	124
4.5 Conclusion	126
CHAPTER: 5 Evaluation of Precision ID Ancestry Panel in Qatar population by next generation sequencing using the Ion Torrent Personal Genome Machine (PGM) ...	127
Part one: Sequencing performance, allele frequencies and other forensic parameters of the panel in the Qatari population.	127
5.1 Introduction	127
Objectives	128
5.2 Materials and Methods.....	128
5.3 Results	128
5.3.1 Sensitivity study	128
5.3.2 Quality of the sequencing runs	132
5.3.3 Sequence Coverage.....	135
5.3.4 Performance of the Precision ID Ancestry Panel	139
5.3.5 Statistical analysis of SNPs data	139
5.4 Discussion.....	144

Part two: Ancestry Inference of the Precision ID Ancestry panel and STRUCTURE analysis in Qatar population	147
5.5 Results	147
5.5.1 Ancestry prediction with the default HID SNP Genotyper	147
5.5.2 Ancestry prediction with a customized HID SNP Genotyper (Addition of Qatar population data)	153
5.5.3 STRUCTURE Analysis	157
5.7 Discussion	159
5.8 Conclusion	161
CHAPTER: 6 Analysis of forensic casework samples by Precision ID Ancestry Panel – Manual and Automated AmpliSeq Workflow	163
6.1 Introduction	163
Objectives	163
6.2 Materials and Methods	164
Part one: Manual Library Preparation Workflow	165
6.3 Results	165
6.3.1 Quality of the sequencing runs	166
6.3.2 Sequence Coverage	169
6.3.3 Performance of the Precision ID Ancestry Panel	173
Part two: Automated Library Preparation Workflow	180
6.4 Results	180
6.4.1 Quality of the sequencing runs	181
6.4.2 Sequence Coverage	183
6.4.3 Performance of the Precision ID Ancestry Panel	186
Part three: Ancestry Inference of the Precision ID Ancestry panel.....	191
6.5 Results	191
6.5.1 Ancestry prediction with the default HID SNP Genotyper	191
6.5.2 Ancestry prediction with the customized HID SNP Genotyper with Qatar population Data	194
6.6 Discussion	198
6.7 Conclusion	204
CHAPTER: 7 General Discussion, conclusion, limitation and future work	205
7.1 General Discussion	205
7.2 Combined results of the Precision ID Identity panel and Precision ID Ancestry Panel.....	206
7.2.1 Population samples results	206
7.2.2 Forensic samples results	207

7.3 Limitations/Challenges of MPS Analysis	214
7.4 Conclusion	214
7.5 Alternative MPS systems.....	216
7.6 Scope for future studies	216
CHAPTER: 8 References.....	218
CHAPTER: 9 APPENDICES	245
9.1 Appendix 1: Map of Qatar	245
9.2 Appendix 2: Ethical approval letter from the University of Central Lancashire’s STEM Ethics Committee.	246
9.3 Appendix 3: Explanation of the Hardy –Weinberg equilibrium, pairwise <i>FST</i> , linkage disequilibrium (LD), and forensic parameters.	247
9.4 Appendix 4: Table describes the Ion Sphere™ particle (ISP) density, summary, and read length quality metrics for the unaligned sequencing reads.	249
9.5 Appendix 5: Data showing Heterozygote balance values as minimum, maximum and median Hb. The lowest median values are highlighted in yellow color.	251
9.6 Appendix 6: Data showing noise level values as minimum, maximum and median of the Precision ID Identity panel. Y SNPs are shaded in grey.	254
9.7 Appendix 7: Data showing allele frequency of 90 autosomal SNPs in Qatar population.	257
9.8 Appendix 8: Data showing forensic parameters of 90 autosomal SNPs of the Precision ID Identity panel for 105 Qatari samples, (PM: probability of match, PD: power of discrimination, PE: Power of Exclusion, PIC: polymorphism information content, and TPI: Typical Paternity Index).	259
9.9 Appendix 9: Data showing observed [Obs.Het] and expected [Exp.Het] heterozygosities and P values from an exact test for HWE across 90 SNPs typed in 105 individuals from three populations. Green represents minimum Obs.Het and orange represents minimum Exp.Het.	262
9.10 Appendix 10: Significant LD detected in P1, P2, P3.	266
9.10.1 Appendix A: Significant LD detected in P1 in yellow color.	266
9.10.2 Appendix B: Significant LD detected in P2 in orange color.....	266
9.10.3 Appendix C: Significant LD detected in P3 in blue color.	266
9.11 Appendix 11: Data showing allele frequency of 165 autosomal SNPs in Qatar population.	270
9.12 Appendix 12: Data showing the Observed (Obs.Het) and expected (Exp.Het) heterozygosities and p-values from an exact test for Hardy-Weinberg Equilibrium (HWE) across 165 SNPs. Cells highlighted with yellow represents *p-value <0.05, blue represents monomorphic SNPs.	274
9.13 Appendix 13. Data showing the forensic parameters of 165 SNPs of The Precision ID Ancestry panel for 300 Qatari samples. PM: match probability, PD: power of discrimination, PE: Power of Exclusion, PIC: polymorphism information	

content and TPI: Typical Paternity Index). Cells highlighted with green color represents the highest discrimination power (PD) whereas cells highlighted with blue color represents the highest power of exclusion (PE).	278
9.14 Appendix 14: Admixture Prediction and Population likelihood values (population samples).	282
9.15 Appendix 15: Data showing the combined results of the 105 Qatari samples from Precision ID Ancestry and Identity panel experiments.	291
9.16 Appendix 16: Data showing the combined results of the 60 casework samples from Precision ID Ancestry and Identity panel experiments.	309
PUBLICATIONS	315
CONFERENCES/ PROCEEDINGS	315

LIST OF ABBREVIATIONS

°C: Degrees Celsius

AIMs: Ancestry Informative Markers

ASMs: Ancestry-Sensitive Markers

ASO: Allele specific oligonucleotide

BC: Before Christ

bp: Base Pair

CE: Capillary electrophoresis

CMOS: Complementary metal-oxide semiconductor

CPD: Combined Power of Discrimination

CPE: Combined Probability of Exclusion

CPM: Combined Power of Match

DTT: Dithiothreitol

EVCs: Externally visible characteristics

FDP: Forensic DNA Phenotyping

Hb: heterozygote balance

HGP: Human Genome Project

HWE: Hardy-Weinberg Equilibrium

IISNPs: Individual identification SNPs

INDEL: insertion- deletion polymorphisms

ISPs: Ion Sphere particles

LD: Linkage Disequilibrium

LISNPs: Lineage Informative SNPs

MLPs: Multi locus probes

MPS: Massive parallel sequencing

NGS: Next generation sequencing

NL: Noise level

NTC: No template control

OLA: Oligonucleotide ligation assay
PC: Promega Corporation
PCI: Phenol Chloroform Isoamyl-alcohol
PD: Power of Discrimination
PE: Power of Exclusion
PIC: Polymorphism Information Content
PISNPs: Phenotype Informative SNPs
PK: Proteinase K
PM: Probability of Match
pmol: Picomoles
SBE: Single-base extension
SBS: Sequencing-by-synthesis
SLPs: Single locus probes
SNP: Single Nucleotide Polymorphism
STR: Short Tandem Repeat
TFS: Thermo Fisher Scientific
TPI: Typical Paternity Index
TSC: The SNP consortium
TSS: Torrent Suite™ Software
YCC: Y chromosome Consortium
Ybp: years before the present

LIST OF FIGURES

Figure 1.1 Diagram showing hybridization with allelic-specific oligonucleotides (ASO) [Taken from (Sobrinho et al., 2005)].	10
Figure 1.2. Flow diagram showing modular design of some of the assays for SNP genotyping. Coloured arrows are used to show the reaction principles, assay format and detection methods that make up a particular genotyping method [Taken from (Syvänen, 2001)].	11
Figure 1.3. Data showing the backbone of the new Y chromosome binary haplogroup tree and its 20 major clades [Taken from (Karafet et al., 2008)].	19
Figure 1.4. Flow diagram showing the library preparation [Taken from (Børsting, 2015)].	24
Figure 1.5. Diagrams illustrating the Emulsion PCR technique. [Taken from (Vierstraete, 2012)].	25
Figure 1.6. Schematic diagram showing the enrichment process. [Taken from (Vierstraete, 2012)].	26
Figure 1.7. Schematic diagram showing cross-section of a single well of an Ion Torrent sequencing chip [Taken from Life Technologies Corporation (2011)].	27
Figure 1.8. Pie charts showing the 165 Precision ID Ancestry Panel [Taken from (Thermo Fisher Scientific, 2016a)].	29
Figure 1.9. Pie charts showing the 124 Precision ID Identity Panel [Taken from (Thermo Fisher Scientific, 2016b)].	29
Figure 1.10. Map showing the State of Qatar and the eight municipalities [The map provided from The Centre for Geographic Information Systems (GIS) in the Ministry of Municipality and Environment]	31
Figure 1.11. Qatar population (in thousands) from 1950-2008. [Taken from (The Permanent Population Committee, 2009)].	33
Figure 1.12. Time course graph showing the Qatari population (in thousands) from 1986-2015. [(Ministry of Development Planning and Statistics, 2016a)].	34
Figure 1.13. Map of State of Qatar showing the distribution of total population by municipality between 2010 and 2015. [Taken from Ministry of Development Planning and Statistics (2015)].	36

Figure 1.14. Labour force (15 years and above), in millions (2014-2018) [Taken from Planning and Statistics Authority, 2019a].	37
Figure 1.15. Bar charts showing the relative distribution of labour Force by nationality, sex and sector, 2018 [Taken from (Planning and Statistics Authority, 2019a)].	38
Figure 3.1 The different panels showing the ISP density percentage and ISP summary for Chip#1, Chip#2, Chip#3 and Chip#4 that was taken from summary report for each chip.	77
Figure 3.2. Data showing the heterozygote balance values (median) for autosomal SNPs.	81
Figure 3.3. Data showing noise level values (median) of the Precision ID Identity panel.	82
Figure 3.4. Column plot of allele frequencies of 90 autosomal SNPs of Precision ID identity panel in Qatar population.	84
Figure 3.5. Plot graph showing the forensic parameters of 90 autosomal SNPS of The Precision ID Identity panel for 105 Qatari samples, (PM: probability of match, PD: power of discrimination, PE: Power of Exclusion, PIC: polymorphism information content, and TPI: Typical paternity index.	85
Figure 3.6. (A) and (B) Bar plots of Observed heterozygosities (Obs.Het) and expected heterozygosities (Exp.Het) across 90 autosomal SNPs of the Precision Identity panel typed in 105 Qatari individuals from P1, P2 and P3 groups.	89
Figure 3.7: A typical example of the information provided for the predicted Y-haplo-group.	94
Figure 4.1. Chip diagrams showing the ISP density percentage and ISP summary for Chip#5 and Chip#6 taken from summary report for each chip.	103
Figure 4.2. (A) and (B) Column chart showing maximum and median of the coverage values for each SNP.(A) Showing the 90-Autosomal SNPs. (B) Showing the 34- YSNPs included in the Precision ID Identity panel resulted from casework samples.	107
Figure 4.3. Diagrams showing (A) the STR profiles of AmpFISTR Identifiler Plus system of S3-BC3-P5 and S4-BC4-P5, (B) partial YSNPs profile collected from the sample, (C) Y haplogroup predictions and (D) RMP results of these samples.	118
Figure 4.4. Diagrams showing (A) part of the electropherogram (epg) generated from S24 and S25 using GlobalFiler, (B) part of the electropherogram (epg) generated from	

S24 and S25 using YFiler Plus, (C) part of the genotype generated from these samples using 124 SNPs included in the Precision ID Identity Panel and (D) Y haplogroup prediction and E) RMP results of these samples.	121
Figure 4.5. Diagrams showing (A) the electropherograms generated from sample S28 using GlobalFiler, (B) part of the partial autosomal SNPs profile, (C) the partial part of the Y-SNPs and (D) Y haplogroup prediction and RMP results of sample S28-BC28-P5.	123
Figure 5.1. The above diagram and table illustrating the coverage bar chart (A) and coverage table (B) for sample 0.01 ng, replicate 2, with 23 loci dropped out. The example in B shows rs1407434 failing the Quality Control; major allele frequency balance is 78.95 which is >65% (MAF).	131
Figure 5.2. The above bar chart diagram (A) and (B) table illustrating the coverage charts and coverage tables, respectively for sample 0.01 ng, replicate 1, with no profile.	132
Figure 5.3. Diagrams showing the three sections that evaluate the quality of the run (run summary).....	133
Figure 5.4 . Images of sequencing chips with highest ISP loading density. (A) Chip#7 QAT181-QAT210). (B) Chip#9 (QAT241-QAT270).	134
Figure 5.5. An image of chip#6 (QAT151-QAT180) with the lowest average ISP density.	135
Figure 5.6. Data showing the Coverage results for sample QAT271 with no drop out. (A) Coverage table and (B) Coverage bar charts.	136
Figure 5.7. Data showing the Coverage results for sample QAT299, which had no genotype. (A) Coverage table and. (B) Coverage bar charts.....	137
Figure 5.8. Column plot of allele frequencies of 165 autosomal SNPs of Precision ID Ancestry panel in Qatar population.	141
Figure 5.9. Bar plots of Observed heterozygosities (Obs.Het) expected heterozygosities (Exp.Het) and p- values across 165 SNPs of the Precision Ancestry panel. Monomorphic SNPs are: rs3811801, rs671 and rs1800414.	142
Figure 5.10. Plot graph showing the forensic parameters of 165 SNPS of The Precision ID Ancestry panel (PM: probability of match, PD: power of discrimination, PE: Power of Exclusion, PIC: polymorphism information content, and TPI: Typical paternity index.	143

Figure 5.11. Maps and data in (A) and (B) show the Admixture Prediction and the Population Likelihoods for sample QAT237.....	148
Figure 5.12. Maps and data in (A) and (B) showing sample QAT299 which had no genotype.	149
Figure 5.13. Maps and data showing re-analysis results of sample QAT255.(A) Admixture Prediction. (B) Population Likelihoods (default) and (C) Qatar Population Likelihoods (addition Qatar data).	156
Figure 5.14. The graphical representation of assigning K value based on the Evanno method generated from STRUCTURE-Harvester software for 8 populations.	157
Figure 5.15. Original images of structure analysis results based on 165 SNPs; the most likely number of clusters (K) is K=4.	158
Figure 5.16. Bar charts showing Geographic distribution for 197 Qatari samples analyzed with default and customized HID SNP Genotyper plugin.....	160
Figure 6.1. Diagrams showing Chips 1-5 displaying the ISP density percentage and ISP summary for Chip#1, Chip#2, Chip#3, Chip#4 and Chip#5 which were taken from summary report for each chip.	168
Figure 6.2. Column charts showing maximum and median of the coverage for each SNP included in the Precision ID Ancestry panel resulted from casework samples (manual library preparation experiment).	172
Figure 6.3. Diagrams showing the ISP density percentage and ISP summary for Chip#1 and Chip#2 which was taken from summary report for each chip.	182
Figure 6.4. Column charts (A) and (B) showing maximum and median of the coverage for each SNP included in the Precision ID Ancestry panel resulted from casework samples (Automation library preparation experiment).....	185
Figure 6.5. Bar charts showing the geographic distribution for 100 casework samples. Data were expressed as a percentage.	191
Figure 6.6. Bar charts showing the geographic distribution for the 47 samples processed with automation workflow. Data expressed as a percentage.....	193
Figure 6.7. Maps and data showing the Admixture Prediction, Population Likelihood and Qatar Population Likelihood result of a typical example number S18-BC18-P5. It was a saliva stain from a cigarette butt with a DNA input 1.9 ng.....	197

Figure 7.1. Bar charts and tabular data showing the Y haplogroup frequency based on 34 -Y-SNPs included in Precision ID Identity with Geo-region as predicted from Precision ID Ancestry panel.....207

Figure 7.2. Pie charts showing the Rodriguez-Flores et al., (2016) pie charts of the Y haplo-group frequencies among the 53 male Qatari divided into Q1 (Beduion), Q2 (Persian-south Asian) and Q3 (African).....211

LIST OF TABLES

Table 1.1. Data showing categories of SNP markers [Taken from (Butler 2011)].	13
Table 1.2. Data showing reads, run time and analysis time for the Ion PGM system for using Precision ID SNP panels [Taken from (Thermo Fisher Scientific, 2017a)].	27
Table 1.3 . Data showing number of samples per run for each chip format. [(Thermo Fisher Scientific, 2017a)].	28
Table 1.4. Table showing Qatar population by place of resident at census night and the gender. [Taken from (Ministry of Development Planning and Statistics, 2016b)].	35
Table 1.5 Population and population density by municipality, 2015 [Taken from (Ministry of Development Planning and Statistics, 2015)].	35
Table 2.1 List of frequently used chemicals and reagents.	42
Table 2.2. Data showing the total number of samples collected from the Ministry of Interior departments and from other volunteers.	43
Table 2.3. Data showing the total number of samples collected from different municipalities; each municipality was coloured according to its colour on the map attached in Appendix 1.	44
Table 2.4. Data showing the total number of casework samples collected from different municipalities.	45
Table 2.5. Data showing the total number of different biological samples collected.	46
Table 2.6. Table showing total number of samples according to their extraction methods.	48
Table 2.7. Data showing the total samples analysed with Precision ID Ancestry Panel and Precision ID Identity Panel.	54
Table 2.8. Data showing the 105 Qatari samples studied with Precision ID Identity Panel.	55
Table 2.9. Data showing the PCR run programme used for the samples with 1 ng.	56
Table 2.10. Data showing the PCR run programme used for the samples (casework samples) with <1 ng.	56
Table 2.11. Data showing partial digestion PCR programme.	57
Table 2.12. Data showing an example of diluted barcode adapter mix for 4 runs.	57

Table 2.13. Data showing the thermal cycler programme for ligation step.....	58
Table 2.14. Data showing the Precision DL8 kit contents.....	60
Table 2.15. Data showing the panel recommendation for the automated library preparation with Ampliseq workflow.	62
Table 2.16. Data showing the appropriate information entered for each run.....	63
Table 2.17. Data showing the analysis parameters for each panel.	64
Table 2.18. Data showing the total number of samples templating on Ion Chef manually and automatically.....	65
Table 3.1. Data in table showing the HID SNP Genotyper Plug-in quality checks.	73
Table 3.2. Data showing the results of the sensitivity study.	74
Table 3.3. Data showing the average mean depth of sequencing for the 4 Chips.....	78
Table 3.4. Data displaying the autosomal SNPs which showed the lowest coverage values among the 105 samples.....	79
Table 3.5. Data showing the YSNPs with the lowest coverage values among the 84 male samples.	79
Table 3.6. Data showing the nine samples showed partial Y-SNP profiles.....	80
Table 3.7. Data showing the three divided populations (groups) from the Qatar population, with a total of 105 samples. The groups were made according to the geographical locations on the map of Qatar.	86
Table 3.8. Data showing the pairwise F_{ST} values between 3 tested population based on 90 Autosomal SNPs included in the panel.	90
Table 3.9. Data showing the significant linkage disequilibrium at $p < 0.05$ in three populations of Qatar.	91
Table 3.10. Data showing the plugin information of the Y Haplogroup predicted [taken from HID SNP Genotyper Plugin (Thermo Fisher Scientific, 2017b)]......	93
Table 3.11. Data showing the total of 10 different Y haplogroups were observed among 84 Qatari individuals.	94
Table 3.12. Data showing the mutation that defined the given Haplogroup.....	95
Table 4.1. Table showing the total number of casework samples collected for the Precision ID Identity panel experiment.	102
Table 4.2. Data showing the average mean depth for Chip#5 and Chip#6.	104

Table 4.3. Table showing the autosomal SNPs which showed the lowest coverage values.	104
Table 4.4. Table showing the Y-SNPs with the lowest coverage values.	105
Table 4.5. Data showing the full SNP profiles were seen in 27 casework samples.	108
Table 4.6. Data showing partial SNP profiles were seen in 28 casework samples.	111
Table 4.7. The results illustrating the 5 samples that failed to generate any profile. ...	113
Table 4.8. Data showing the Y- haplogroup obtained form 43 casework samples.	114
Table 4.9. Table showing 17 casework samples reported with "No haplo-groups found".	114
Table 5.1. Data showing the results of the sensitivity study. Confidence level is with reference to the ancestry prediction.	130
Table 5.2. Data showing the Average ISP density for each chip used. The lowest value is highlighted.	134
Table 5.3. Data showing the average mean depth obtained from 10 chips. The highlighted row represents the Chip with the lowest coverage.....	136
Table 5.4. Data showing the SNPs with the highest coverage among the 299 samples.	138
Table 5.5. Data showing the SNPs with the lowest coverage among the 299 samples.	138
Table 5.6. The data showing the overall genotype for the 300 samples in each chip with chip loading%.	139
Table 5.7. Data showing the comparison of 165 autosomal SNPs in the Precision ID Ancestry Panel data between Qatar population and other world populations.	146
Table 5.8. Data showing the most likely population of origin results predicted of 299 samples from Qatari population were collected from eight municipalities. Each colour corresponds to the color of the municipality presented in Qatar map found the Appendix 1.	151
Table 5.9. Data showing the most likely population of origin predicted for the 299 individuals.	152
Table 5.10. Data showing the predicted populations for the Qatari individuals studied with Precision ID Ancestry Panel. Column in the middle represents the previous results. The right column represents the results of the modified plug-in with Qatar population	

likelihood. The green line shows the number of individuals predicted as Qatari. Orange indicates populations with reduced numbers assigned; grey represents no change. .154

Table 6.1. Data showing the total number of casework samples were used in manual library preparation workflow.....165

Table 6.2. Data showing the average ISP density summarized for each chip used.....166

Table 6.3. Table showing the mean depth for the 5 Chips used in this study.169

Table 6.4. Data showing the SNPs with the lowest coverage.....170

Table 6.5. Data showing the full SNP profiles were seen in 47 casework samples. Columns shaded brown contained suboptimal DNA template.173

Table 6.6. Results showing the partial SNP profiles were seen in 53 casework samples.176

Table 6.7. Data showing the 11 samples that failed to generate any profile.....179

Table 6.8. Data showing the different types of samples were used in automation library workflow.180

Table 6.9. Data generated from 48 samples following automated library preparation. Sixteen samples out of 48 were included in both workflow (manual and automation workflow) and are highlighted in grey colour. Yellow cells represent the samples that were amplified using 22 cycles whereas blue cells represent the sample that were amplified for 25 cycles (these are slightly different than the recommended usage for the manual workflow, which advise 21 and 26 cycles).....181

Table 6.10. Data showing the average mean depth for Chip#1 and Chipe#2 used in automation workflow.183

Table 6.11. Data showing the SNPs with the lowest coverage values.....183

Table 6.12. Data showing the full SNP profiles which were seen in 29 samples. Samples in grey were previously examined using the manual workflow.186

Table 6.13. Data showing the partial SNP profiles were seen in 18 samples, rows highlighted in blue were already studied with manual library workflow.....189

Table 6.14. Data showing the Population Likelihood results from the most likely population of origin of the 100 DNA profiles.....192

Table 6.15. Data showing the Population Likelihood results. The most likely population of origin of the 47 DNA profiles.193

Table 6.16. Data showing the re-analysis with Qatari Population Likelihood results with the most likely population of origin of the 100 DNA profiles (Manual Workflow).	195
Table 6.17. Data showing the re-analysis with Qatar Population Likelihood results; the most likely population of origin of the 47 DNA profiles (Automation Workflow).....	196
Table 6.18. Data showing the profile quality of the 16 samples studied in both workflow: Manual and Automation library workflow.	199
Table 6.19. Data showing the SNPs with the lowest coverage with the two workflows.	200
Table 6.20. Results showing the SNPs with the lowest coverage values seen in two experiments.	200
Table 6.21 Data showing the predicted populations for the Qatari population and casework samples studied with Precision ID Ancestry Panel. (Grey cells represent the population seen in both experiments, yellow only in population samples and green cells only in casework samples).	202
Table 7.1. Data showing the Y-chromosome haplogroup frequency observed for Qatar collected from the present (N= 84 Qatari) and previous study (N= 72 Qatari).....	210

ACKNOWLEDGEMENT

In the name of Allah, the Merciful "Allah will raise up in ranks those who believed among you and those who have been given knowledge. Allah is Aware of what you do". Surat Al-Mujadilah Ayat 11.

This PhD would not have been possible to do without the moral support, guidance, experiences and cooperation that I received from many people.

First, of all I am greatly thankful to the Ministry of Interior for their constant moral and material support and cooperation for giving me this opportunity to carry out my research abroad and to have my practical experiments in the Forensic DNA and Biology laboratory in Doha - Qatar.

I would like to express my sincere gratitude to my supervisor Dr William Goodwin who has provided me valuable advice, academic stimulus, support and kindness during my whole PhD project. I would like to thank Dr Arati Iyengar, Dr Sibti Hadi and Professor Jaipaul Singh for their help and guidance.

My special appreciation and thanks go to my local supervisor Brigadier Ibrahim Abdulla Al-Bin Ali with his trust, invaluable guidance, support and generous help to achieve all the goals of the PhD journey.

I wish to express my deep thanks and heartfelt gratitude to my parents, the ones who can never ever be thanked enough, for the overwhelming love and care and they remain a constant source of blessing for me and their prayer enables me to achieve my higher education and a lot in this world. My father, who has always discussed, examined, and guided me with his insightful advice. My mother, whoever was my companion on my academic journey, her prayers that blessed my days. My sincere thanks to my family especially my sister Eida and my brothers Shaheen, Mesfer and Ali for their love, understanding and constant support over the years. Big thanks extended to my cousins Mashael Ahmed Al-Mohandi and to all of my family members for their love and encouragement, it will always mean the most to me.

Last but not least, many thanks to all my beloved friends at UCLan and I am especially grateful to Aisha Ahmed Al-Baker and Sajda Hassan Karmallah at Qatar Forensic Lab for their constant help and moral support. To Dr Satya Sudheer Pydi, Ashraf Badwi and Matt Phipps (from Thermo Fisher Scientific) for their invaluable technical support.

CHAPTER: 1 Introduction

1.1 A historical perspective of forensic science and its application

The ABO blood grouping system described by Karl Landsteiner (Landsteiner 1900) was the first genetic marker used in forensic analysis to distinguish between individuals. This simple test was useful in case of exclusion but was difficult when an inclusion was made (Butler, 2005). The serological techniques were extended with the discovery and the use of several blood groups markers and soluble blood serum protein markers. They were analysed in combination to produce high discrimination results. However, there were limitations with the use of serological tests such as the low discrimination power and, in many cases, the large amount of biological sample that was required. Also, the proteins are not present at sufficient levels for typing in many types of evidence, and they are unstable in biological samples affected by environmental conditions (Budowle & Van Daal, 2008; Geserick & Wirth, 2012; Goodwin et al., 2011).

In 1984, Alec Jeffreys discovered hypervariable minisatellite loci (Jeffreys et al., 1985a). These minisatellites loci were detected by hybridization of multi locus probes (MLPs) to Southern blots of restriction-enzyme-digested genomic DNA. MLPs were used successfully in paternity testing and immigration cases but were difficult to interpret and were replaced with single locus probes (SLPs) detection assays, which produced simplified patterns that were easier to interpret. In 1983, Kary Mullis developed the polymerase chain reaction (PCR) technique (Mullis et al., 1986). The introduction of PCR revolutionized molecular biology, including forensic DNA. With the development of PCR forensic DNA analysis became more rapid and sensitive, allowing minute amounts of degraded DNA to be analysed (Jobling & Gill, 2004; Parson, 2018; Tamaki & Jeffreys, 2005).

1.2 Recent DNA typing in forensic analysis

DNA typing has played an important role in forensic analysis since its introduction in 1985 (Jeffreys et al., 1985a). The first use of DNA fingerprinting was to resolve an immigration dispute in 1985 (Jeffreys et al., 1985b); it was soon after used to link the two murders of young girls in Leicestershire to Colin Pitchfork who was convicted of both murders in 1988 (Seton, 1988). Since then, variable tandem repeat (VNTR) markers have been replaced by short tandem repeat (STR) markers. STR with several developments leading to a robust system for analysing biological evidence (Jobling & Gill, 2004; Shrivastava et al., 2016).

STRs are short repeated sequences (2–6 bp), and the number of repeat units is highly variable among individuals, which lead to the high discrimination power. STRs are short when compared to VNTRs and therefore easier to amplify using PCR. In addition, amplifying multiple STR loci, which can then be detected in one assay using capillary electrophoresis, enables the generation of DNA profiles with a high power of discrimination (Wyner et al., 2020; Butler, 2007; Roewer, 2013).

STR typing has become the primary tool for individual identification in criminal casework, paternity testing, and identification of missing persons (Parson et al., 2016). Currently, forensic DNA casework worldwide depends on the use of STR technology with over 20 loci routinely analysed in a single PCR. The commercial companies Thermo Fisher Scientific (TFS), Promega Corporation (PC) and Qiagen provide widely-used STR kits with amplicons that range in size from 100 base pairs (bp) to 450 bp for crime scene casework, paternity and database samples.

Since the early 2000s, two kits, the Identifiler™ and PowerPlex® 16 system, were widely used. More powerful commercial kits are now available, such as an improved Identifiler™ released as Identifiler™ Plus, to improve the typing of degraded and inhibited forensic samples (Ip et al., 2014). Subsequently released kits added more loci and therefore had higher powers of discrimination were also more sensitive; examples of widely used kits with 20 plus loci include Globalfiler™ (TFS) (Martín et al., 2014), PowerPlex® Fusion 6C (PC) and Investigator® 24plex QS (Qiagen™) (Zgonjanin et al., 2017).

An insertion- deletion polymorphisms (INDEL) kit has also become commercially available; the Investigator DIPplex® kit (Qiagen™) is an assay for 30 bi-allelic INDEL markers and amelogenin (LaRue et al., 2012).

In addition to the above robust autosomal STR kits, several Y-STR multiplexes available and are routinely used in certain forensic casework analysis (Corach et al., 2001; Hanson & Ballantyne, 2007). Lineage Y-chromosome markers in combination with autosomal STRs can provide additional data to paternity disputes of male offspring, and also to male identification cases involving human skeletal remains such as in mass disaster victim and missing person identification where only distant relatives are available (Kayser, 2017; Ambers et al., 2018). Y-STR profiles can also play an important role when analysing mixtures samples female- male samples and help in the identification of persons in sexual assault cases. For Y chromosome analysis the latest kits include PowerPlex® Y23 (PC) (Oostdik et al., 2014) and YFiler™ Plus with 27 loci (TFS) (Gopinath et al., 2016).

Typing forensic samples can be challenging, leading to incomplete profiles or no profiles with the current STR commercial kits. Especially when dealing with old and/or degraded DNA. Also, sometimes the analysis of STRs is limited in close relative kinship testing and crime cases with no suspect. Looking to a complementary tool(s) to use with STR typing may help in some contexts (Canturk et al., 2014).

Recent advances in next generation sequencing, also called massive parallel sequencing (MPS), has started to be used in forensic genetics, with the aim of providing more discriminating data from both routine and challenging samples. The MPS approach offers the potential to enhance profiling by amplifying shorter amplicons and providing greater sensitivity and higher success rates with degraded material. New MPS platforms have been released and new 'forensic ready' SNP panels and commercial kits have been developed for SNP sequencing, STR sequencing and mitochondrial sequencing (Guo et al., 2016; Iozzi et al., 2015).

1.3 Single nucleotide polymorphisms (SNPs)

SNPs are defined by single base differences at a specific position in the genome and represent about 90% of human genetic variation. For such, a single base change to be

considered a SNP, the minor frequent allele should have a frequency of $\geq 1\%$ (rarer SNPs are characterized, but are of limited utility for forensic applications). SNPs are abundant in human genome and exist at approximately 1 in every 1000 bp. SNPs in coding and regulatory regions can affect the encoded protein and have been the focus of medical, pharmaco-genetics and forensic phenotyping. SNPs that occur in the non-coding region are typically selected for studies in population genetics, evolutionary studies and forensic studies (Carracedo, 2005; Collins et al., 1998; Daniel & Walsh, 2006; Venter et al., 2001).

The simplest difference between two homologous DNA sequences is a base substitution where one base is exchanged for another: transitions are more common than transversions. Transition is when a pyrimidine base is exchanged for another pyrimidine (for example, C for T), or a purine for another purine (for example, A for G), When a pyrimidine is exchanged for a purine, or *vice versa*, this is a transversion (Jobling et. al., 2013).

Within the category of SNPs an insertion or deletion (INDEL) of a single base is also a possibility but the formation and analysis of these single nucleotide indels is different from the transversions and transitions. An INDEL is defined as insertion or deletion of bases in the genome, it is the result of either insertion or deletion of a DNA sequence ranging from one nucleotide to hundreds of nucleotides (Butler, 2011).

SNPs are usually bi-allelic (or di-allelic); the nucleotide that is present is one of two possibilities (Daniel & Walsh, 2006). Multiple-allele SNPs are potentially useful for forensic DNA analysis as they can provide more discrimination power than normal binary SNPs. Tri-allelic SNPs are much less frequent than their binary equivalents, while tetra-allelic SNPs with all four A, C, G and T alleles at one nucleotide position will be even more rare (Phillips et al., 2015). Tri-allelic SNPs are loci with three substitutions recorded at the variant nucleotide position. The six genotypes in tri-allelic SNPs clearly increase the overall level of polymorphism per marker compared to binary SNPs, making tri-allelic variants compelling markers for forensic application, since a multiplex of such SNPs. Moreover, they allow the analysis and interpretation of samples with simple mixture by detection of the third alleles. The allele frequency of tri-allelic SNP markers tends to show high levels of population stratification, making many tri-allelic

SNPs suitable for use as ancestry markers (Phillips, 2012; Phillips et al., 2020). Two studies have identified tri-allelic SNPs for forensic casework purposes (Phillips et al., 2004; Westen et al., 2009). The study of Phillips et al. (2015) identified 24 tetra-allelic SNPs with good discrimination power in Europeans or Africans, although few loci were informative for East Asians. In addition to being SNPs very useful tool in forensic fields, SNPs also provide powerful tools for disease diagnosis and drug response studies. In disease genetics studies utilising SNPs, the aim is to identify SNPs that cause changes in cellular biological processes inducing diseased states ((reviewed by (Kim & Misra, 2007)) (Toma et al., 2018; Enoch et al., 2006; Araújo et al., 2016; Bhaskar et al., 2017; Wen et al., 2016; Bandera et al., 2013).

1.3.1 SNP Resources and Databases

1.3.1.1 The Human Genome Project HGP

The Human Genome Project (HGP) was initiated in 1990 in the US with the goal of determining the complete nucleotide sequence of the human genome, to identify and map the thousands of human genes and more so to make these data accessible to all researchers for further scientific study. It coordinated by the National Institutes of Health (NIH) and the U.S. Department of Energy (DOE), and completed the first draft in 2001. Celera Genomics also participated to assemble the human genome. The International Human Genome Sequencing Consortium used hierarchical shotgun sequencing (HS) in contrast to Celera Genomics that adopted the whole-genome shotgun approach. HS is based on the sequencing of overlapping bacterial artificial chromosomes (BACs) with known locations in the human genome and provides a guaranteed route for producing an accurate finished genome sequence, because the sequence assembly is anchored to the genome. In WGS approach the genome was fragmented into individual random reads, which was then stitched together to produce a single contig for each chromosome.

In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. In 2003, HGP leaders announced the Human Genome Project was completed, and then in October 2004 it was published the near complete human genome sequence (Human Genome

Sequencing Consortium, 2004; Chial, 2008; Lander et al., 2001; Waterston et al., 2002; Venter et al., 2001; Giani et al., 2020). Throughout, the HGP was instrumental in facilitating and driving the development of high-throughput technologies for sequencing DNA using the Sanger method, which is usually referred to as first-generation sequencing (Hood & Rowen, 2013).

1.3.1.2 The SNP consortium (TSC)

The SNP consortium (TSC) was established in 1999 as a collaboration between major pharmaceutical companies, the Wellcome Trust, and five leading academic and genome sequencing centres (<http://snp.cshl.org>) with the mandate to produce a public resource of SNPs with an initial target to discover 300,000 SNPs in two years. The TSC project ended with several million human SNP markers mapped and characterized, in samples from African, European and Asian populations (Thorisson & Stein, 2003).

1.3.1.3 HapMap Project

In 2002, the International HapMap Consortium developed the International HapMap Project to build a haplotype map (“HapMap”) of the human genome. The aim was to identify patterns of genetic variation and linkage disequilibrium (LD) in 269 individuals from four populations with ancestry in Africa, Europe, and Asia (Belmont et al., 2003). The project was conducted in three phases. Phase I released in 2005, with 1,007,329 SNPs identified. Phase II released in 2007 with over 3.1 million SNPs genotyped from the same individuals. In phase three the sample number increased to 1,184 from 11 global populations, released in 2010. The HapMap 3 data was released with 1.6 million common SNPs genotyped using Affymetrix 6.0 and Illumina 1.0 Million SNP mass arrays (Belmont et al., 2003; Belmont et al., 2005; Frazer et al., 2007; International HapMap 3 Consortium, 2010).

1.3.1.4 HGDP-CEPH Diversity Panel Database

The HGDP-CEPH Human Genome Diversity Cell Line Panel database was designed to receive and store polymorphic marker genotypes, copy number variant (CNVs) calls, and Sanger DNA sequences which are generated by users of the DNA samples provided by the HGDP-CEPH Diversity Panel. The data are publicly accessible via <http://www.cephb.fr/hgdp/>. The panel was developed by a group of scientists at

Stanford University that collaborated with the Centre for the Study of Human Polymorphism (CEPH) at the Foundation Jean Dausset in, Paris. Its collection consists of 1,064 lymphoblastoid cell lines (LCLs) from 52 populations around the world. LCLs were collected for reasons of accuracy and renewability (Cavalli-Sforza, 2005). The populations are from Africa, Europe, the Middle East, South and Central Asia, East Asia, Oceania and America. The samples have been characterised at more than 650,000 SNP loci using the Illumina BeadStation. The resource has been a useful resource for SNP discovery and characterisation, and for determining global sequence variation at various loci (Cann et al., 2002; Cavalli-Sforza, 2005; Fullerton & Lee, 2011; Li, J et al., 2008; Rosenberg et al., 2002) even though political, ethical, legal and social implications have impeded the project since its inception (Greely, 2001). Reference population of HGDP-CEPH were also used in several studies such as 'MAPlex' (Multiplex for the Asia-Pacific) that aimed to improve differentiation of East Asian, South Asian and Near Oceanian populations (Phillips et al., 2019) and Eurasiaplex which is a SNP panel for differentiating European and South Asian ancestries (Phillips et al., 2013).

1.3.1.5 The Single Nucleotide Polymorphism Database (dbSNP)

This is a public-domain archive for simple genetic polymorphisms. The National Center for Biotechnology Information (NCBI) developed the Single Nucleotide Polymorphism Database dbSNPS (<https://www.ncbi.nlm.nih.gov/snp/>) to supplement GenBank in collaboration with the National Human Genome Research Institute (NHGRI). It includes SNPs, INDELS, retro-transposable element insertions and STRs (microsatellites).

In 2008, there were approximately 12.8 million human reference SNPs (<https://www.ncbi.nlm.nih.gov/snp/>). dbSNP is integrated with other public variation databases, such as The SNP Consortium (TSC) where all SNPs listed in the SNP consortium are listed to dbSNP (Smigielski et al., 2000).

1.3.1.6 1000 Genomes Project

The 1000 genomes project was initiated in 2008 to establish a deep catalogue of human genetic variation that could serve as a baseline for further research into the relationship between genotype and phenotype and for identifying the genetic basis of human disease (Zheng-Bradley & Flicek, 2017).

The main goal was to find and characterise most genetic variants that have frequencies of at least 1% and identify the haplotypes and characterize their LD patterns.

The pilot phase evaluated different strategies for genome-wide sequencing with high throughput platforms including Illumina, Roche 454 and SOLiD from Life Technologies through three projects and the samples from the extended HapMap project.

A trio project: whole-genome shotgun sequencing at high coverage (average 42x) of two families (one Yoruba from Ibadan, Nigeria (YRI); one of European ancestry in Utah (CEU)), each including two parents and one daughter.

Low-coverage project: whole-genome shotgun sequencing at low coverage (2–6x) of 59 unrelated African individuals (YRI), 60 unrelated European individuals (CEU) and 60 unrelated East Asian individuals 30 unrelated Han Chinese individuals in Beijing (CHB) and 30 unrelated Japanese individuals in Tokyo (JPT).

Exon project: targeted high-coverage (average >50x) sequencing of 8140 exons from 906 randomly selected genes (total of 1.4 Mb), in 697 individuals from seven populations of African (YRI, LWK), European (CEU, TSI), and East Asian (CHB, JPT, CHD) ancestry (1000 Genomes Project Consortium, 2010).

After the pilot phase, Phase 1 of the project focused on low coverage (2–6x) whole genome sequence data and exome sequence data (50–100x) for 1092 individuals from 14 populations from Europe, East Asia, sub-Saharan Africa and the Americas. Phase 1 identified 38 million single nucleotide polymorphisms, 1.4 million INDELS, and more than 14,000 larger deletions (1000 Genomes Project Consortium, 2012). Phase 2 focussed on data production and technical development (Zheng-Bradley & Flicek, 2017).

In 2015 the final phase (Phase 3) of the 1000 Genomes was published and includes data for 2,504 individuals from 26 populations in Africa (AFR), East Asia (EAS), Europe (EUR), South Asia (SAS), and the Americas (AMR). Phase 3 discovered 88 million variants (84.7 million SNPs, 3.6 million short INDELS, and 60,000 structural variants) (1000 Genomes Project Consortium, 2015).

1.4 Typing and detection methods

1.4.1 SNP typing methods

Several methods of SNP typing assays have been emerged, based on different methods of allelic discrimination with different detection platforms. SNPs assays are based on the following mechanisms: allele specific hybridization, primer extension, oligonucleotide ligation, and invasive cleavage. (Sobrino et al., 2005; Twyman, 2009).

1.4.1.1 Allele specific hybridization

Hybridization using allele specific oligonucleotide (ASO) probes is a relatively simple method for discriminating between alleles: two ASO probes are needed, one specific for each allele and with the polymorphism located in a central portion of the probe sequence. When optimized only the perfectly matched probe-target hybrids are stable and remain bound to the target sequence as shown in Figure 1.1. ASO probes with reverse dot-blot formats were the first method employed to detect SNP polymorphisms in a forensic context (Sobrino et al., 2005; Twyman, 2009).

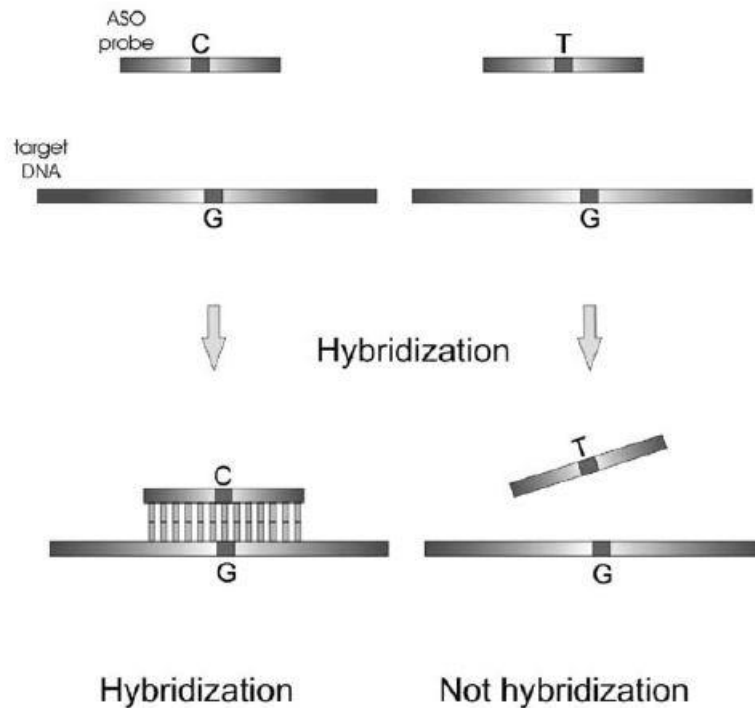


Figure 1.1 Diagram showing hybridization with allelic-specific oligonucleotides (ASO) [Taken from (Sobrino et al., 2005)].

1.4.1.2 Primer Extension

1.4.1.2.1 Mini-sequencing

Single-base extension (SBE) technique is also known as mini-sequencing. This method involves a primer that anneals to its target DNA immediately adjacent to the polymorphic SNP. Extension with a DNA polymerase adds a single labelled nucleotide to the probe (primer). SNaPshot is a commercial kit for single-base extension released by Life Technologies (TFS) and incorporated fluorescently labelled probes followed by capillary electrophoresis. Over the last decades the mini-sequencing has been the most widely used SNP detection method in forensic laboratories, especially the SNaPshot, largely because it utilised the same instrumentation as used in STRs analysis (Sobrino et al., 2005), until the introduction of Massive Parallel Sequencing (MPS)-based systems. Several panels have been developed in different forensic laboratories using SNaPshot technology (Fondevila et al., 2013; Lao et al., 2010; Phillips et al., 2007a; Sanchez et al., 2006).

1.4.1.2.2 Pyrosequencing

Pyrosequencing is a sequencing-by-synthesis technology and is a real-time sequencing process. During DNA sequencing when dNTPs are incorporated into a nascent DNA strand, an inorganic pyrophosphate P_{PPi} is released, leading to an emission of a light that is detected by CCD camera (Harrington et al., 2013; Sobrino et al., 2005). In 2005, the first “massively parallel” “next-generation sequencer” based on the pyrosequencing reaction was developed by Rothberg and colleagues: the Roche 454 sequencer (Harrington et al., 2013).

Invasive cleavage and allele specific oligonucleotide ligation are other methods, but not particularly suitable for forensic applications because of the large amount of DNA required for the analysis (Phillips et al., 2007b; Tobler et al., 2005; Tomas et al., 2008).

1.4.2 Detection methods

The product of each type of method can be analysed using different detection methods such as fluorescence, luminescence, and mass measurement (Kwok, 2001; Sobrino et al., 2005; Twyman, 2009) as shown in (Figure 1.2).

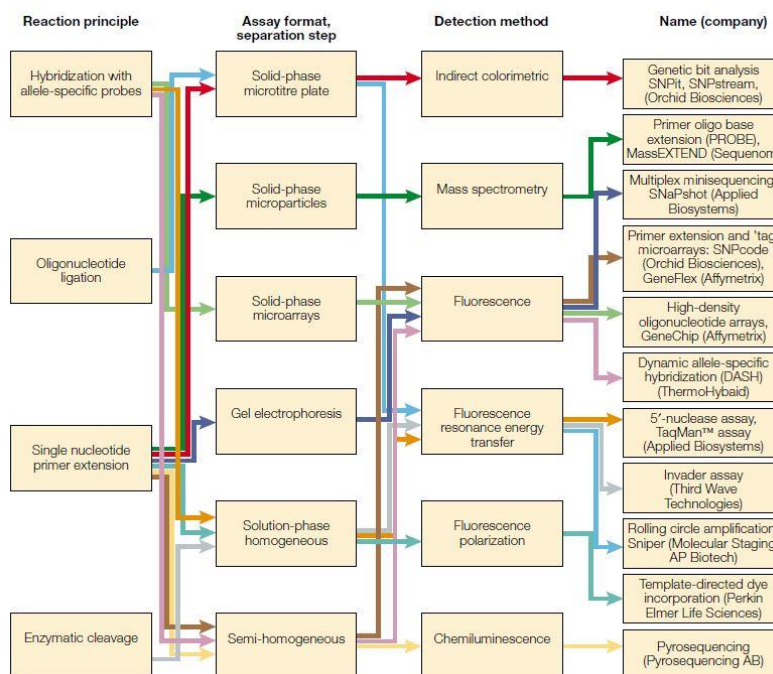


Figure 1.2. Flow diagram showing modular design of some of the assays for SNP genotyping. Coloured arrows are used to show the reaction principles, assay format and detection methods that make up a particular genotyping method [Taken from (Syvänen, 2001)].

Until recently, the only platform that was widely used in forensic genetics was the SNaPshot system, as it used capillary electrophoresis systems. However, with the advent of MPS and in particular the Illumina and Thermo Fisher platforms typing large numbers of SNPs is now possible for a large number of forensic laboratories.

1.5 SNPs in forensic analysis

A panel discussion on SNPs and their application in forensic identity and relationship testing was held at the International Society of Forensic Genetics (ISFG) in Copenhagen, Denmark in 2007. Scientists presented their SNP panels, discussing the benefits, limitations and requirements for SNP analysis in terms of both implementation and interpretation (Budowle, B. & Van Daal, 2008; Butler et al., 2008; Butler, 2011). The applications were classified into four general uses as shown in Table 1.1.

Table 1.1. Data showing categories of SNP markers [Taken from (Butler 2011)].

Categories	Characteristics	Examples
Identity SNPs: Individual identification SNPs (IISNPs)	SNPs that collectively give very low probabilities of two individuals having the same multi-locus genotype.	FSS 21plex (Dixon et al., 2005) SNPforID 52plex(Sanchez et al., 2006) Kidd group SNPs(Pakstis et al., 2010)
Lineage SNPs: Lineage Informative SNPs (LISNPs)	Sets of tightly linked SNPs that function as multi-allelic markers that can serve to identify relatives with higher probabilities than simple bi-allelic SNPs.	mtDNA coding region SNPs (Coble et al., 2004) Japanese Y-SNPs (Mizuno et al., 2010) Haplotype blocks (Ge et al., 2010)
Ancestry SNPs: Ancestry Informative SNPs (AISNPs)	SNPs that collectively give a high probability of an individual's ancestry being from one part of the world or being derived from two or more areas of the world	SNPforID 34-plex (Phillips et al., 2007a) 24 SNPs (Lao et al., 2010) FSS YSNPs (Wetton et al., 2005)
Phenotype SNPs: Phenotype Informative SNPs (PISNPs)	SNPs that provide a high probability that the individual has particular phenotypes, such as a particular skin colour, hair colour, eye colour, etc.	Red Hair (Grimes et al., 2001) "Golden" gene pigmentation (Lamason et al., 2005) IrisPlex eye colour (Walsh et al., 2011a)

1.5.1 Identity SNPs

Forensic laboratories analyse most casework samples using the available commercial STR kits, mini-STRs or in some cases both used. However, some cases/samples remain challenging, and the use of additional markers may be useful in cases where DNA is highly degraded and also in complex kinship cases. Samples can be typed with SNPs which are small in length 50-150 bp. SNPs selected for identity should ideally have low F_{ST} values (i.e., low variation between populations) and high heterozygosity (Budowle & Van Daal, 2008). Several efforts have been made to build SNP multiplex assays.

1.5.1.1 SNPforID 52plex

In 2003, several members of the European forensic DNA typing community launched a SNPforID project for developing SNP loci for forensic DNA analysis. Two sets of SNPs, identity and ancestry, in highly multiplexed assays using unlinked loci that were spread throughout the human genome were developed. Population data were also gathered to measure SNP allele frequencies in various groups (Butler, 2005; Butler, 2011). In addition, a web-based browser was developed to support the assay and included a world map which is clickable at any population selected. The SNPforID studies were carried out on populations from across Africa, America, Europe, East-Asia, Central-South Asia, Middle East and Oceania (Amigo et al 2008).

The identity multiplex was a set of 52 SNPs amplified in a single tube PCR reaction followed by two single-base extension (SBE) reactions and visualized with electrophoresis (CE) system, 3 SNPs were later removed from the assay due to their poor performance (Børsting et al., 2009a). The amplicon size ranged from 59 bp to 115 bp and showed full profiles from degraded samples which where gave partial STR profiles (Sanchez et al., 2006; Børsting et al., 2009a).

1.5.1.2 Ken Kidd SNPs

For Individual identification Kidd et al. (2006) designed a preliminary panel of 19 SNPs with high heterozygosity and low F_{ST} values and gives an average match probability of $<10^{-7}$ in most of the 40 populations studied; 432 SNPs were screened. An expanded panel was subsequently developed with 92 SNPs for individual identification (Kidd et al., 2006; Pakstis et al., 2007). The 92 IISNPs showed an average heterozygosity >0.4 with F_{ST} values <0.06 based on the global 44 populations studied. This was later reduced to a set of 86 SNPs of the 92 IISNPs, none of which showed significant linkage disequilibrium (LD) in any of the 44 populations. The remaining 6 IISNPs show strong LD in most of the 44 populations. From the 86 SNPs 45 showed no genetic linkage and were proposed to be of value for the testing of close biological relationships. Match probabilities in the 44 tested populations ranged between 10^{-31} and 10^{-35} (Kidd et al., 2012; Pakstis et al., 2010).

1.5.2 Ancestry SNPs

Estimating the source origin of an unknown sample can potentially provide intelligence when the routine identity STR/SNP genotypes do not match to an individual by identifying the likely population of origin (Phillips et al., 2007a). Ancestry Informative Markers (AIMs) have alleles with frequency differences between populations; ideally a SNP would be fixed in one population and at a low frequency or absent in another population. SNPs tend to be better at estimating biogeographical origins than STRs because SNPs are more stable than STRs (Butler, 2011).

1.5.2.1 Early panels of ancestry informative markers (AIMs)

Several panel of ancestry informative markers (AIMs) were been developed for estimating ancestry (Daniel & Walsh, 2006). In 2003 Shriver and co-workers listed a panel of 34 AIMs based on the relationships of skin pigmentation to individual ancestry. The study was undertaken in two populations with primarily African ancestry, African Americans and African Caribbean. Significant correlations between estimates of individual ancestry and skin pigmentation were observed.

An ancestry classifier was built by Frudakis and co-workers in 2003 with 211 SNPs that were screened from panels of human pigmentation and xenobiotic metabolism genes. A panel with 56 SNPs was identified with varying allele frequencies between groups of unrelated donors of Asian, African and European descent. An algorithm incorporating the 56 SNP was 99%, 98%, and 100% accurate for inferring individuals of European, African, and Asian origin.

A further panel was published in 2008 that comprised 176 autosomal AIMs. the AIMs were identified in two stages by screening publicly available databases: the first stage identified 71 AIMs and the second identified another 105 AIMs. The panel was able to differentiate individual biogeographical ancestry (I-BGA) and admixture proportion from four continental ancestral population: European, West Africans, Indigenous Americans, and East Asians (Halder et al., 2008) . Single-base primer extension was used in a tagged fluorescent assay using the 25 K SNP stream ultra-high-throughput genotyping system (Beckman Coulter, Fullerton, CA) (Halder et al., 2008). The panel led to a commercially available multiplex for biogeographical ancestry known as "DNAWitness" by a company DNAPrint (DNAPrint, Inc., Sarasota, FL).

1.5.2.2 SNPforID 34-plex

A single tube assay was developed with 34 SNPs for forensic ancestry tests and based on the most informative AIMs that showed highly contrasting allele frequency distributions between the three major population-groups (Phillips et al., 2007a). Training sets for the classifications were created for each population from collecting two population samples from sub-Saharan Africans, Europeans, and East Asians. To test the classification performance the CEPH human genome diversity cell line panel (CHPH-HGDP and achieved consistently high classification probabilities, even when a reduced subset of SNPs was used). A single SNP rs727811 was later replaced with the highly informative rs3827760 that shows a near-fixed East Asian specific allele and also some adjustments were done on amplification and CE chemistry. The revised SNP set was examined using Standard Reference Material (SRM) 2391c and were typed along with the standard forensic positive control DNAs: AB/Promega 9947a; Qiagen XY5, and; Promega 2800M and 66 reference populations from the 1000 Genomes Project and the CEPH panel (Fondevila et al., 2013). The 34-plex was found to be useful to analyse highly degraded samples and routine forensic casework samples (Phillips et al., 2012).

1.5.2.3 Ancestry-Sensitive Markers ASMs

A panel comprising with 47 SNPs was developed by (Kersbergen et al., 2009) and served as a panel of ASMs at a continental level. In a study by Kersbergen and co-workers they preferred to use the term ancestry-sensitive markers (ASMs) because, in their opinion, ancestry sensitivity better reflects the uncertainties related to such marker, whereas ancestry informativity implies that they clearly reveal ancestry. They analysed 74 worldwide Y Chromosomal Consortium (YCC) samples from six geographical regions with the Affymetrix Mapping 10K assay. STRUCTURE software was used to detect genetically distinct subgroups. Later, in order to identify a single best performing set of ASMs they used a pairwise F_{ST} ranking procedure among all pairs of genetic subgroups.

1.5.2.4 Lao panels

Lao et al., (2006) showed a set of reduced 10 SNPs was able for detecting continental population structure. An algorithm was to identify a set of markers that maximized the genetic differentiation between populations and minimized the number of markers needed. Analysis of the 10 SNPs set on samples from the CEPH Human Genome Diversity Panel (HGDP-CEPH) was able to assign the genetic ancestry of individuals from all four continents represented in the original data set (Lao et al., 2006). Lao et al., (2010) also developed a panel of 24 SNPs that were genotyped by two 12-plex SNaPshot multiplex reactions.

1.5.2.5 Seldin group 128-AIMs

Kosoy et al. (2009) validated and developed a panel of SNPs that comprised 128 AIMs for estimating continental ancestry. They analysed 825 individuals from 20 designated populations; TaqMan assays were used to type the alleles. They found that with 128 AIMs and subsets as small as 24 AIMs to be useful to confirm the origin of subjects from particular continents, and to correct for population stratification in admixed population sample sets (Kosoy et al., 2009).

1.5.2.6 41-SNP panel

Nievergelt et al., (2013) developed a 41-AIM AISNP panel for multiplex genotyping for the ABI SNPLEX genotyping system and also a subset with 31 AIMs for analysis using the Sequenom iPLEX system. The panel was found to be highly informative for estimating ancestry with samples from Africa, the Middle East, Europe, Central/South Asia, East Asia, the Americas and Oceania. They validated the 41-panel with a large reference data with 4018 unrelated subjects from 120 global populations (Nievergelt et al., 2013).

1.5.2.7 Kidd Laboratory 55 AIMs

Kidd et al., (2014) developed and published a panel of reduced highly informative SNPs that showed seven to eight biogeographic regions could be distinguished. To develop the panel, they analysed 3884 individuals from 73 populations using TaqMan assays (Kidd et al., 2014).

1.5.3 Lineage SNPs

SNPs locate on mtDNA and Y-Chromosome play useful role in forensic DNA testing particularly in providing individual's lineage due lack of recombination and a low mutation rate. These markers can be informative for evolutionary and Kinship studies (Budowle & Van Daal, 2008).

The Y chromosome represents only 2% of the human genome with an overall size of approximately 60 Mb. The majority of the Y chromosome comprises the largest non-recombining block of DNA (NRY) 95% (Hammer & Zegura, 2002; Quintana-Murci et al., 2001). Single nucleotide polymorphism (SNP) and STR haplotypes on the NRY provide powerful tools for inferring the histories of populations (Mountain et al., 2002). The simplicity of the SNP markers that they are likely to have two allelic and also their much lower mutation rates (about 2×10^{-8}) geographic ancestry signatures are retained for longer. Y-SNP s are informative for paternal bio-geographic ancestry inference, population evolution and migration patterns (Jobling, 2001; Muro et al., 2011).

1.5.3.1 Y-haplogroup and YCC tree

Y-haplogroups can infer population membership of a male lineage that may contribute to forensic casework analysis (Lessig et al., 2005; Muro et al., 2011). The term haplogroup indicates the combination of alleles at multiple SNPs, while the combination of alleles at multiple Y-STRs on a single Y chromosome defines a Y-STR haplotype (Hammer & Redd, 2006; Y Chromosome Consortium, 2002).

The Y chromosome Consortium (YCC) was an international group of scientists lead by Michael Hammer from the University of Arizona (Butler, 2011). In 2002 the YCC constructed a highly resolved tree of NRY binary haplo-groups by genotyping most published PCR-based markers on a common set of samples. Based on the tree it created a new nomenclature system that was flexible enough to allow later modifications (Y Chromosome Consortium, 2002). In 2002, the YCC published phylogenetic tree of 153 different haplo-groups as 18 major clades based on almost 250 markers genotyped. In 2003 the haplo-group tree was slightly modified and updated (Jobling & Tyler-Smith, 2003), and then in 2008 was revised and expanded to 311 haplo-groups with 2 new major clades (S and T), combining approximately 600

binary markers (see Figure 1.2 ; (Karafet et al., 2008; Y Chromosome Consortium, 2002).

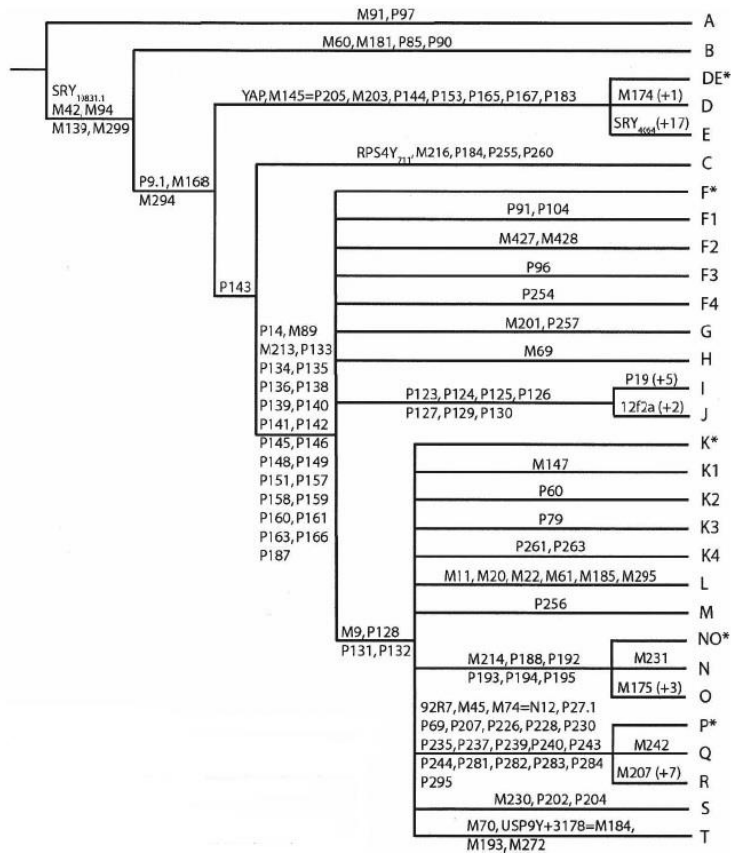


Figure 1.3. Data showing the backbone of the new Y chromosome binary haplogroup tree and its 20 major clades [Taken from (Karafet et al., 2008)].

Mutations labelled with the prefix “M” (standing for “mutation”) were published by Underhill et al. (2000, 2001). Many of the mutations with the prefix “P” (standing for “polymorphism”) were described by Hammer et al. in 1998 and 2001 (Hammer & Zegura, 2002; Karafet et al., 2008; Muro et al., 2011; Y Chromosome Consortium, 2002).

1.5.4 Phenotype SNPs

In forensic analysis the term Forensic DNA Phenotyping (FDP) is the prediction of externally visible characteristics (EVCs) from DNA. Predicting the physical appearance of an individual from crime scene evidence can assist the criminal investigations by reducing the number of potential suspects in cases (Walsh et al., 2013). Phenotypic SNP predictions focus on hair, skin, and eye pigmentation. Eye colour depends on the

amount and distribution of melanin pigment in the iris. A single-tube Iris-plex test was developed to predict eye colour (Canturk et al., 2014; Walsh et al., 2011b). A second system was developed in 2013 by Walsh et al., called “Hirisplex” test a combined hair and eye colour prediction (Canturk et al., 2014). HirisPlex-S DNA test system (S for skin) for the simultaneous prediction of eye, hair, and skin colour. The results from the three predictor systems IrisPlex, HirisPlex, and HirisPlex-S were combined into a publicly available interactive tool <https://hirisplex.erasmusmc.nl/> (Chaitanya et al., 2018; Marano & Fridman, 2019). In 2017, the VISible Attributes through GENomics (VISAGE) Consortium was created and aimed to use MPS techniques to generate more information about the appearance of unknown persons. It was composed of members of Academia, Police, and Justice Institutions from eight European countries, with the goal to enable facial composites from DNA evidence (Marano & Fridman, 2019).

1.6 SNP Typing Technologies

1.6.1 First generation Sequencing

In 1970s, the DNA sequencing witnessed the release of two methods by Sanger and colleagues and Maxam and Gilbert by using chain termination and fragmentation techniques. Sanger and Maxam-Gilbert sequencing technologies were considered as the First Generation Sequencing Technology (Kchouk et al., 2017; van Dijk et al., 2014).

1.6.2 Next generation sequencing (NGS)

NGS is also referred to massively parallel sequencing (MPS) that sequence the DNA in massively parallel way with high coverage and high throughput of specified targets. MPS is often called next generation sequencing, to differentiate the new technology developments from previous DNA sequencing systems. MPS has some major differences in comparison to Sanger sequencing. Firstly, NGS technologies can sequence millions of DNA strands parallel in a single run and the time required to create GigaBase sized reads is only a few hours. Secondly, when preparing multiple copies for sequencing, the library preparation is in a cell free system removing the need for bacterial cloning of DNA molecules. Thirdly, the sequencing output data is detected directly without the use of electrophoresis (Kchouk et al., 2017; van Dijk et al., 2014).

Several massively parallel sequencing platforms have been developed and become available in the last couple of years.

Roche/454 in 2005 released the first NGS sequencer, the Roche 454. It used pyrosequencing technology and it typically generated around 200,000 reads (~20 Mb per run) of 100 bp. 454 GS FLX Titanium System was released in 2008, the read length was improved and reached to 700 bp with an accuracy reached 99.9% and output 0.7 Gb of data per run. In late 2009, Roche simplified the library preparation and data processing that enhanced the capacity to 14 Gb per run. The major drawbacks of Roche 454 in most cases were the high error rates in homopolymer repeats, low throughput and the high reagent cost. This technology has been discontinued for several years now. Also, signals with too high or too low intensity led to under or overestimation of the number of nucleotides (Liu, L. et al., 2012; van Dijk et al., 2014; Kchouk et al., 2017).

In 2007 Applied Biosystem (Life Technologies) launched the SOLiD system, which sequence by Oligo Ligation Detection (Yang et al., 2014). SOLiD had an accuracy of 99.85% per run. The output data was 3 Gb per run with a read length of 35 bp. Applied Biosystems released the SOLiD 5500xl in late 2010 in which the read length improved to 85 bp, the output reached to 30 Gb per run and the accuracy 99.99% (Bruijns et al., 2018; Liu, L. et al., 2012). SOLiD has high accuracy because each base is read twice, however, read lengths are short and run times long. The errors of sequencing in this technology can be introduced during the ligation cycle (Kchouk et al., 2017).

In 2006, Solexa released Genome Analyzer (GA) then in 2007 Illumina obtained Solexa. It sequences by synthesis (SBS). The data generated by the sequencer was 1 Gb per run. In 2009 improvement had been made to the sequencer and the output reached 85 Gb. In 2010 Life Technology (Thermo Fisher Scientific) introduced a new sequencer in the form of benchtop Ion Personal Genome Machine (PGM) (Reuter et al., 2015). It uses semi-conductor technology to detect nucleotide incorporation into the growing DNA chains. The significant advantage of this technology there is no requirement for optical scanning and fluorescent nucleotides. Fast run times and takes only a few hours. On the other hand, the major disadvantage is the difficulty of interpreting the

homopolymer sequences (more than six bp), which causes insertion and deletion (indel) errors with a rate of about ~1% (Kchouk et al., 2017; Ballard et al., 2020).

Illumina released HiSeq with the capacity to generate 200 Gb of sequencing data, which it subsequently improved to 600 Gb. Then another benchtop was developed in 2011 which is run on the same technology as HiSeq 2000 and has the capability to generate 1.5 Gb of sequence data (Liu, L. et al., 2012). In early 2015, Illumina launched a complete MPS workflow for forensic DNA analysis by the MiSeq FGx™ Forensic Genomics System. The system consists of four components: the ForenSeq™ DNA Signature Prep Kit, the MiSeq FGx™ Reagent Kit, the MiSeq FGx™ sequencing instrument, and the ForenSeq™ Universal Analysis Software (Guo et al., 2017; Jäger et al., 2017). The ForenSeq DNA Signature Prep Kit (Verogen) contains approximately 200 genetic markers in a single test as two primers A and B which includes STRs (autosomal, X and Y STR) and SNPs (identity, ancestry, and phenotypic SNPs). The MiSeq is designed as a fast, personal benchtop sequencer, with run times as low as 4 h and outputs intended for targeted sequencing of small panels.

In 2015, the newest models of Ion Torrent sequencer, the Ion S5 and Ion S5 XL, were released that provided a simplified workflow and fast sequencing through using load-and-go reagents compared to Ion PGM. The sequencers use pre-made reagents and cleaning and wash bottles and not need to use a water with specific conductivity to prepare the cleaning buffers and also don't require a gas supply as with the Ion PGM (Jeffrey & Peter, 2016; Thermo Fisher Scientific, 2017a).

A wide variety of different NGS platforms have been made available. There are some limitations and reasons for forensic labs not to implement the NGS workflow, such as the cost of commercial panels, run time and the time consumed during the library preparation. However, PGM, MiSeq and S5/S5 XL are considered the standard sequencers that have been adopted in the forensic community. Several ready to use panels that facilitate the SNP and STR typing to forensic scientists have been developed and are commercially available (Bruijns et al., 2018).

1.6.3 The Ion torrent PGM system

In 2010, Ion Torrent (now Life Technologies) released the Personal Genome Machine (PGM). The PGM system was developed by Jonathan Rothberg, (see 1.4.1.2.2) the founder of 454 (reviewed by (van Dijk et al., 2014)). The Ion PGM uses sequencing by synthesis (SBS); sequencing by detection of hydrogen ions and this is based on detection of hydrogen ion liberated on incorporation of each nucleotide. Ion semiconductor sequencing measures pH changes induced by the release of hydrogen ions (Ambardar et al., 2016; Kulski, 2016; Reuter et al., 2015).

The combination of Ion AmpliSeq target technology, an Ion PGM system and Ion Chef System provides a simple, fast and high level of automation system to build up template and loading into chips.

1.6.3.1 Chemistry and workflow technology

1.6.3.1.1 Library preparation

The samples are amplified with Precision ID panel primers. The DNA libraries are constructed from the Precision ID Library Kit by using 21 to 26 PCR cycles according to the amount of DNA input. The resulting amplified products are treated with an enzyme to partially digest the amplification primers followed by a ligation step to ligate specific adapters and Ion Express barcodes to the ends of the amplicons. The adapters or the PCR primer tags can have specific sequences for clonal amplification of the library, target sequences for the NGS reaction, a key sequence with 4–8 nucleotides used for quality control of the NGS reaction and a 6–10 nucleotide barcode for identification of the sample. Before template preparation barcoded libraries are purified and quantified and multiple libraries typically pooled in equal amounts before clonal amplification. The number of samples that can be analyzed simultaneously depends on: the number of available barcodes, the sequencing capacity of the NGS platform, the numbers and sizes of targeted regions and the required sequencing/read depth (Figure 1.4) (Børsting & Morling, 2015).

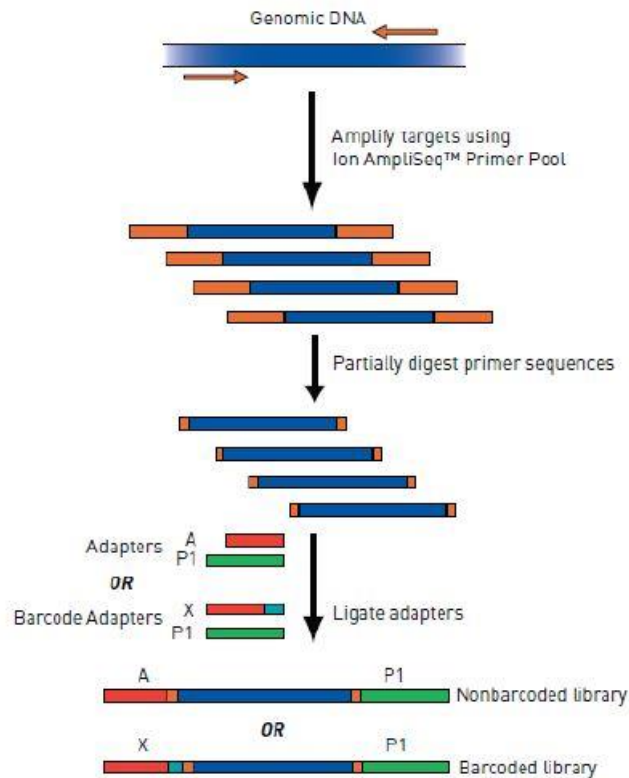


Figure 1.4. Flow diagram showing the library preparation [Taken from (Børsting, 2015)].

1.6.3.1.2 Clonal amplification

After the library preparation the fragments are clonally amplified on Ion Sphere particles (ISPs). An oil–water emulsion is created to create small reaction vesicles (Nakano et al., 2003). Individual DNA molecules are hybridized to a primer on a solid surface, and each molecule is then amplified by PCR in a reaction. Ideally, each vesicle/droplet contains one sphere, one single-stranded template molecule, one of the primers bound to the sphere, and all other reagents necessary for the PCR reaction. Two primers are complementary to the sequence library adapters: one is present in the solution and the other primer is attached to the ion sphere. During the emulsion PCR amplification steps millions of individual DNA molecules get amplified and producing thousands copies of the same template sequence that are bound to the beads (Figure 1.5) (Børsting & Morling, 2015; Buermans & den Dunnen, 2014). The emulsion PCR can be performed in either the Ion OneTouch 2 System or Ion Chef instruments.

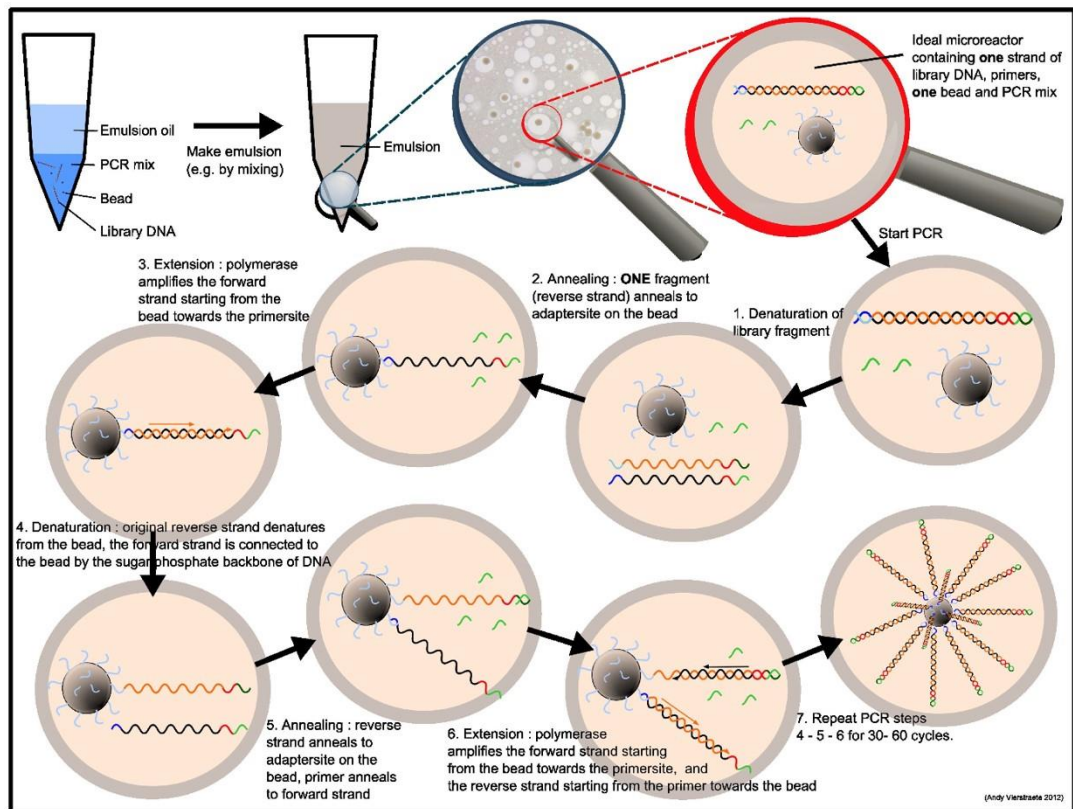


Figure 1.5. Diagrams illustrating the Emulsion PCR technique. [Taken from (Vierstraete, 2012)].

1.6.3.1.3 Enrichment

Enrichment step refers to a wash step for the ISPs. After the emPCR the clonal amplification, ideally, each ISP contains multiple copies of the same DNA fragment. The template ISPs are separated from the non-template ISPs through enrichment process with streptavidin coated beads (Dynabeads MyOne Streptavidin C1 Beads) which can be carried out with either the OneTouch 2™ or Ion Chef Instruments (TFS). The unenriched ISPs are mixed with streptavidin beads, the DNA library fragments are bounded to ISPs on one side while the other side of the fragments are biotinylated. The biotinylated (biotinylated X primer) side will bind to magnetic beads that are coated with Streptavidin allowing them to be collected on a magnet and then all empty ISPs that do not contain any DNA fragment are washed away (Figure 1.6).

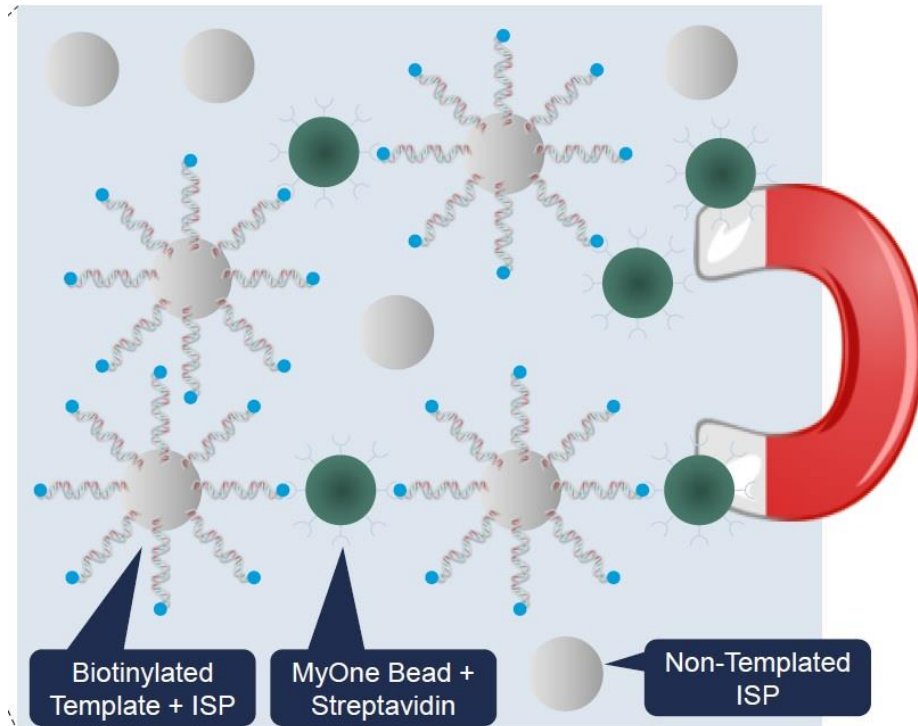


Figure 1.6. Schematic diagram showing the enrichment process. [Taken from (Vierstraete, 2012)].

1.6.3.1.4 Sequencing

Then ISPs coated with template are applied to a sequencing ion chip that holds million copies of DNA. The chip consists of a set of micro wells and the beads are distributed into each well and each micro well has a bead with multiple identical fragments. The detection of incorporated bases is based on the release of a hydrogen ion. During the sequencing step a flow of nucleotides are added one after another T-A-C-G. If the nucleotide is complementary to the DNA strand in the chip well a nucleotide is incorporated which leads to the release of a hydrogen ion that in turn leads to a change in the pH of the solution in that well, which is detected by the ion sensor that translates the chemical change to digital information. The voltage is directly proportional to the number of nucleotides incorporated in Figure 1.7. The output data are analysed using Torrent Server Data Analysis (Buermans & den Dunnen, 2014; Kchouk et al., 2017; Reuter et al., 2015).

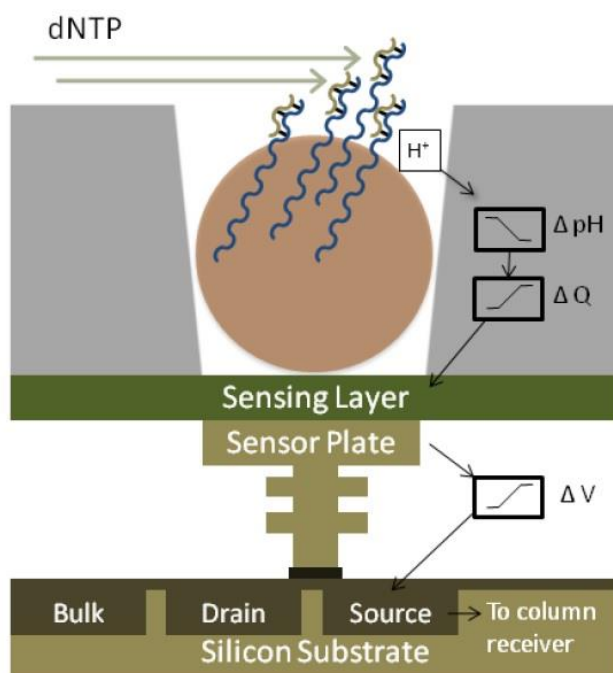


Figure 1.7. Schematic diagram showing cross-section of a single well of an Ion Torrent sequencing chip [Taken from Life Technologies Corporation (2011)].

1.6.3.1.5 Ion sequencing chips

There are three different types of chips for the PGM: 314, 316 and 318 and two for the Ion S5/S5 XL systems which are 520 and 530. The chips differ in the number of wells (Tables 1.2 and 1.3).

Table 1.2. Data showing reads, run time and analysis time for the Ion PGM system for using Precision ID SNP panels [Taken from (Thermo Fisher Scientific, 2017a)].

Types of Chips	Reads	Run Times	Analysis times (Aligned BAM)
314	400,000-550,000	2.3 h	1.5 h
316	2-3 million	3 h	2.5 h
318	4-4.5 million	4.5 h	4 h

Table 1.3 . Data showing number of samples per run for each chip format. [(Thermo Fisher Scientific, 2017a)].

Category	Panel	Ion S5 Chips		Ion PGM Chips		
		Ion 520 Chip	Ion 530 Chip	Ion 314 Chip	Ion 316 Chip	Ion 318 Chip
SNP	Precision ID Ancestry Panel	48	192	4	24	48
	Precision ID Identity Panel	64	264	6	32	64

The Ion torrent workflow can be carried out in line with one of the instruments that facilitate and speed up the workflow. In 2016 Thermo Fisher Scientific released the Ion Chef system that provides a simple and fast workflow for the Ion Torrent sequencers through automated library preparation, automated template preparation and chip loading (Wang et al., 2018).

1.6.3.2 Panels for forensic studies

Illumina and ThermoFisher Scientific have both developed MPS tools and kits. Illumina have designed the Miseq FGX forensic Genomic System specifically for forensic analysis. The ForenSeq™ DNA Signature Prep Kit includes two different primer mixes for library preparation, Primer mix A (PMA) serves mainly for human identification and contains 58 STRs (27 autosomal, 7 X- and 24 Y-chromosome), amelogenin and 94 iiSNPs (identity informative SNPs). Primer mix B (PMB) includes all targets present in PMA plus additional 56 aiSNPs (ancestry informative SNPs) and 22 piSNPs (phenotypic informative SNPs) (Köcher et al., 2018).

In 2014, ThermoFisher Scientific released two ready-to-use forensic panels for human identification applications to work with the Ion PGM, Ion S5 and Ion S5 XL. The HID-Ion AmpliSeq™ Ancestry Panel consists of 165 autosomal SNPs, 55 SNPs from the Kidd panel (Kidd et al., 2014) and 123 SNPs from the Seldin Panel (Kosoy et al., 2009) (Figure 1.8). The HID-Ion AmpliSeq™ Identity Panel includes 90 individual identification SNPs (IISNPs) and 34 lineage informative SNPs (LISNPs) on the Y chromosome (Figure 1.9). Three chips are designed for low throughput (Ion 314 chip), moderate throughput (Ion

316 Chip) and high throughput (Ion 318 Chip) (Børsting & Morling, 2015; Guo et al., 2016).

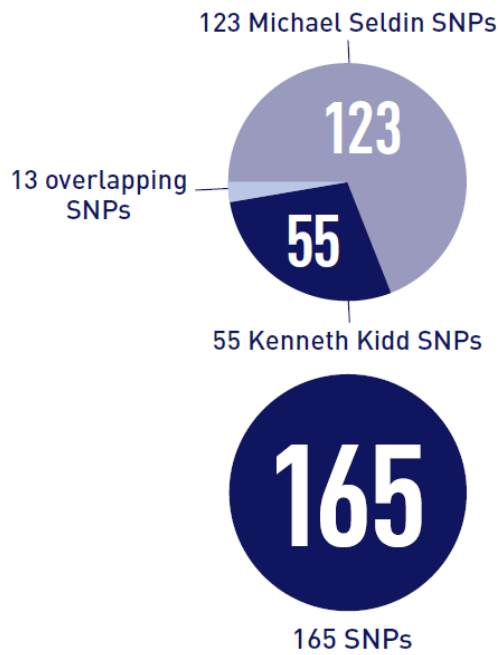


Figure 1.8. Pie charts showing the 165 Precision ID Ancestry Panel [Taken from (Thermo Fisher Scientific, 2016a)].

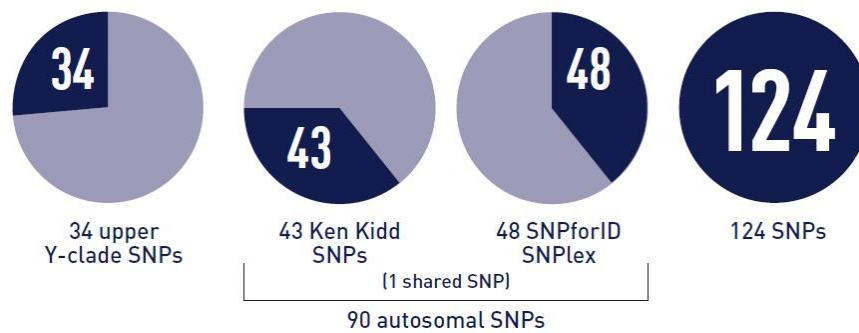


Figure 1.9. Pie charts showing the 124 Precision ID Identity Panel [Taken from (Thermo Fisher Scientific, 2016b)].

Other panels have also been developed by TFS including the Precision ID mtDNA Whole Genome Panel, Precision ID mtDNA Control Region Panel, Precision ID GlobalFiler NGS STR Panels. Promega Corporation has developed the PowerSeq system prototype Auto/Y multiplex to facilitate forensic laboratories to take advantages of MPS. The PowerSeq 46 GY a PCR-based 46-loci multiplex kit, includes the same loci included in PowerPlex Fusion (22 autosomal STR loci, AMEL, and 1 Y-STR locus (DYS391)), as well as the 23 Y-STR loci contained in the PowerPlex Y23 System. This has been designed to run on the Illumina MiSeq FGx system (Riman et al., 2020). Qiagen have developed the forensic identification SNP PCR multiplex Qiagen SNP-ID kit 140 SNP panel, which was firstly studied with Ion PGM and then with Illumina MiSeq (de la Puente et al., 2017).

1.7 Project background

1.7.1 Geography and population distribution of State of Qatar

The State of Qatar is a peninsula with a set of islands situated in the continent of Asia and Doha is the capital of Qatar. It is one of the Arab Gulf countries and from the South has land borders and sea borders with Kingdom of Saudi Arabia and also has sea borders with the United Arab of Emirates, Kingdom of Bahrain and the Islamic Republic of Iran. Qatar occupies around 11,651 sq.km (Figure 1.10) (Ministry of Development Planning and Statistics, 2015).

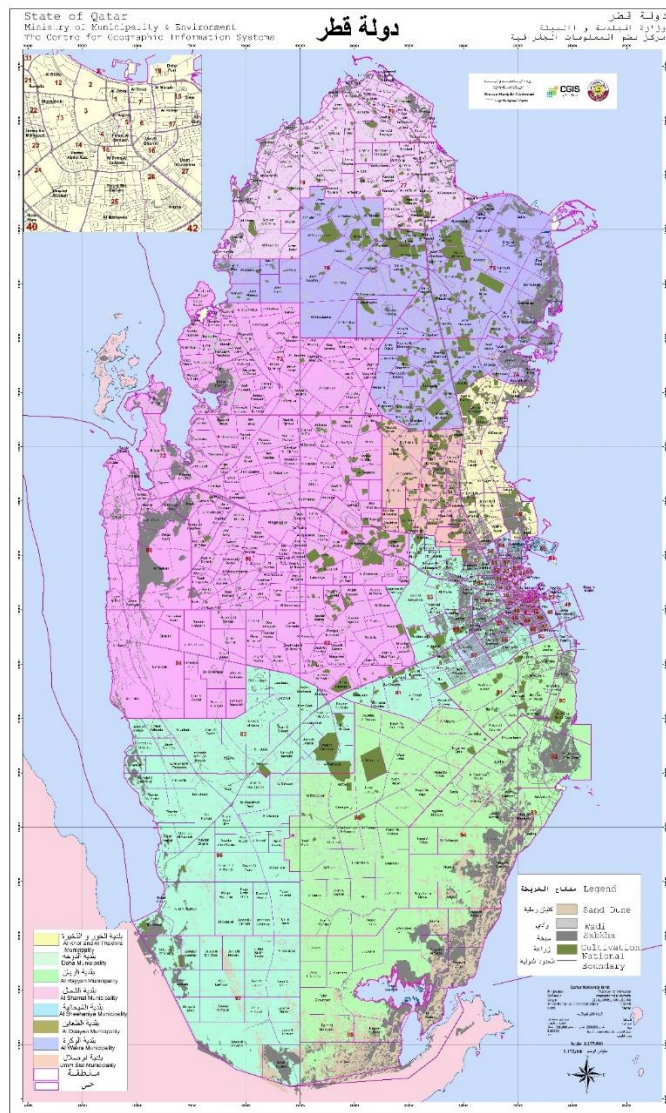


Figure 1.10. Map showing the State of Qatar and the eight municipalities [The map provided from The Centre for Geographic Information Systems (GIS) in the Ministry of Municipality and Environment]

Human habitation of the Qatar Peninsula dates far back to nearly four thousand years BC according to archaeological evidence (Ministry of Foreign Affairs, 2019). The Qatari population society formed through human waves that came to the State of Qatar from the Arabian Peninsula across the Qatari mainland and the eastern coast of the Arabian Gulf (Obaidan, 1982). Qatar's location was the main reason for Qatar to witness the seasonal migrations of Arab tribes across it from the Arabian Peninsula and particularly from the Nejd desert. Also, the Persian Gulf area found commercial prosperity when the ancient Mediterranean flourished with many civilizations (Pérez-Miranda et al., 2006).

According to the Article No.1 of Law No. 38 of 2005, on the acquisition of Qatari nationality 38 / 2005 the following shall be deemed to be Qatari Nationals: “1) Those residents of Qatar who have been resident in the country since before 1930 and who maintained regular legal residence in the country until the enforcement date of the aforementioned Law No. 2 of 1961, 2) Any person who is proved to be of Qatari descent, albeit in the absence of the conditions set forth in the preceding sub-article, and additionally, any person in respect to whom an Emiri decree has been promulgated, 3) Persons to whom Qatari nationality has been reinstated in accordance with the provisions of law and 4) Any person born in Qatar or in a foreign country to a Qatari father in accordance with the preceding Articles”.

The central aim of Qatar National Vision 2030 (QNV 2030) which was launched in 2008 with the aim of transforming Qatar into an advanced country by 2030, “capable of sustaining its own development and provide a high standard of living for all of its people for generations to come” (Planning and Statistics Authority, 2019b). Qatar has experienced a large influx of migrant workers, and foreign workers represent a very large percentage of the total population and constitute the majority of the labour force (Diop et al., 2012).

The population of Qatar subjected to rapid increases over the past four decades. The discovery of oil and gas, as well as the evolution of industries and related services, accelerated development and population growth. Figure 1.11 shows a rapid increase in the people residing in Qatar from 484,000 in 1986 and 522,000 in 1997, to 744,000 in 2004 and more than 1,500,000 in late 2008 (The Permanent Population Committee, 2009).

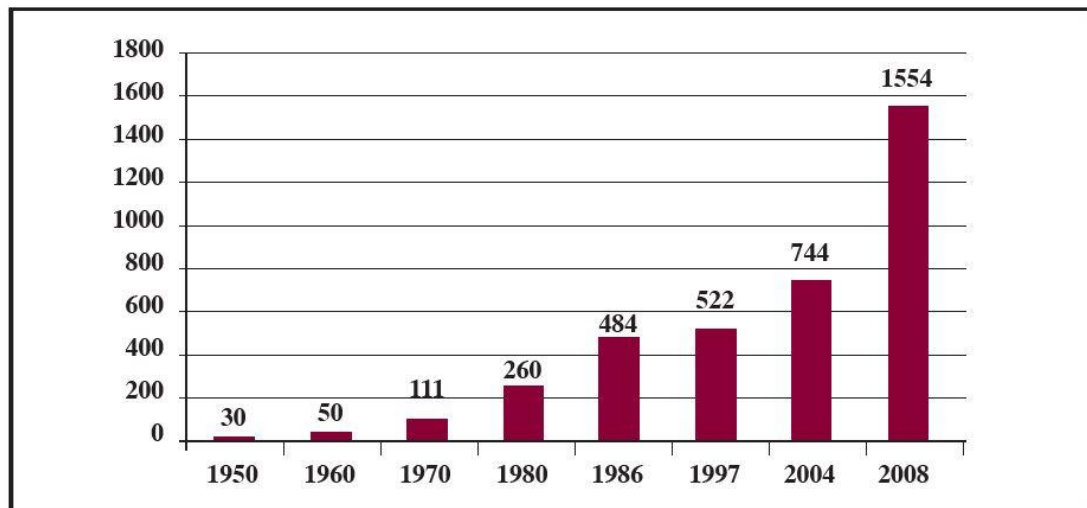


Figure 1.11. Qatar population (in thousands) from 1950-2008. [Taken from (The Permanent Population Committee, 2009)].

Figure 1.12 shows that the population grew from 1,699,000 to 2,405,000 during 2010 to 2015; an increase of 41.5%. Prior to this the population rose 6.5 times during a 29-year period from 1986 to 2015, with an average annual growth rate of 6.7%. The rapid increase in residents, largely driven by foreign workers, is due to the development plans of Qatar National Vision 2030 (Ministry of Development Planning and Statistics, 2016a).

In Qatar the number of live births reached 26,816 in 2016 indicating an increase of 0.7% compared to live births records in 2015. A continuous increase was observed during the period (2007 – 2016) from 15,681 in 2007 to 26,816 live births in 2016 with an annual growth rate of 6% (Ministry of Development Planning and Statistics, 2017).

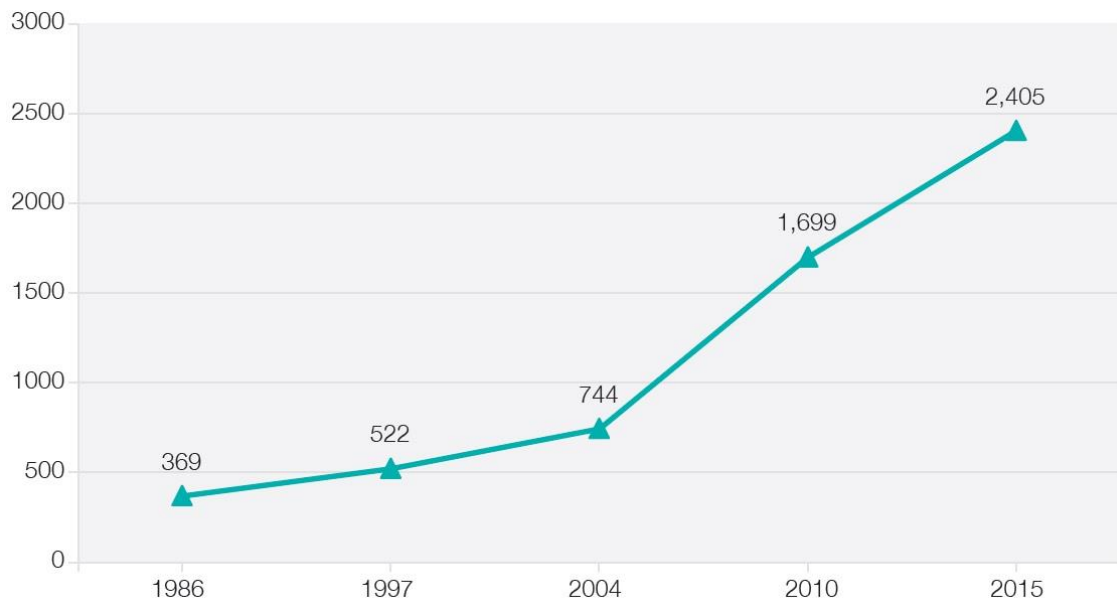


Figure 1.12. Time course graph showing the Qatari population (in thousands) from 1986-2015. [(Ministry of Development Planning and Statistics, 2016a)].

According to the population and social statistics report published in the web site of Ministry of Development Planning and Statistics, the Qatari population is estimated as 2,404,776 as of the April 2015 census – the majority of the residents are non-Qatari, but the exact figure is not published. Fakhro et al., (2016) reported that the current population of the state of Qatar includes more than 1.7 million expatriates primarily from Middle East and North Africa (MENA) and South Asia who have arrived in recent decades and ~ 300,000 Qataris.

The population of Qatar is distributed over eight Municipalities which are Doha, Al Rayyan, Umm Salal, Al Daayen, Al Khor and Al Thakira, Al Shamal, Al Wakra, and Al Sheehaniya. The largest segment of the population resides in Doha and Al Rayyan municipalities while the lowest population density is in Al Shamal municipality (Tables 1.4 and 1.5). Figure 1.12 illustrates the distribution of total population by municipality between 2010 and 2015 (see Appendix 1 Map of State of Qatar showing the location of eight municipalities) (Ministry of Development Planning and Statistics, 2015; Ministry of Development Planning and Statistics, 2016b).

Table 1.4. Table showing Qatar population by place of resident at census night and the gender. [Taken from (Ministry of Development Planning and Statistics, 2016b)].

Places of Residency on Census Night	General total			Public Housing			Labour Camps			Households		
	Total	Females	Males	Total	Females	Males	Total	Females	Males	Total	Females	Males
Doha	956,457	250,027	706,430	1,496	137	1,359	579,349	64,470	514,879	375,612	185,420	190,192
Al Rayyan	605,712	198,929	406,783	3,158	0	3,158	244,135	10,646	233,489	358,419	188,283	170,136
Al Wakrah	299,037	50,934	248,103	1,029	0	1,029	218,922	10,435	208,487	79,086	40,499	38,587
Umm Slal	90,835	30,140	60,695	344	0	344	40,022	1,629	38,393	50,469	28,511	21,958
Al Khor	202,031	21,031	181,000	98	0	98	168,026	3,506	164,520	33,907	17,525	16,382
Al Shamal	8,794	2,271	6,523	0	0	0	4,691	57	4,634	4,103	2,214	1,889
Al Dayyan	54,339	18,294	36,045	0	0	0	24,276	686	23,590	30,063	17,608	12,455
Al Shahniaa	187,571	16,169	171,402	0	0	0	163,461	4,421	159,040	24,110	11,748	12,362
Total	2,404,776	587,795	1,816,981	6,125	137	5,988	1,442,882	95,850	1,347,032	955,769	491,808	463,961

Table 1.5 Population and population density by municipality, 2015 [Taken from (Ministry of Development Planning and Statistics, 2015)].

Municipalities	Areas (Sq km.)	Population	%	Density (per sq.km.)
Doha	219.7	956 457	39.8	4 353.5
Al Rayyan	2 450.1	605 712	25.2	247.2
Al Wakra	2 577.6	299 037	12.4	116.0
Umm Salal	318.4	90 835	3.8	285.3
Al Khor	1 602.2	202 031	8.4	126.1
Al Shamal	859.9	8 794	0.4	10.2
Al Da'ayen	290.2	54 339	2.3	187.2
Al Sheehaniya	3 308.9	187 571	7.8	56.7
Qatar	11 627.0	2 404 776	100.0	206.8

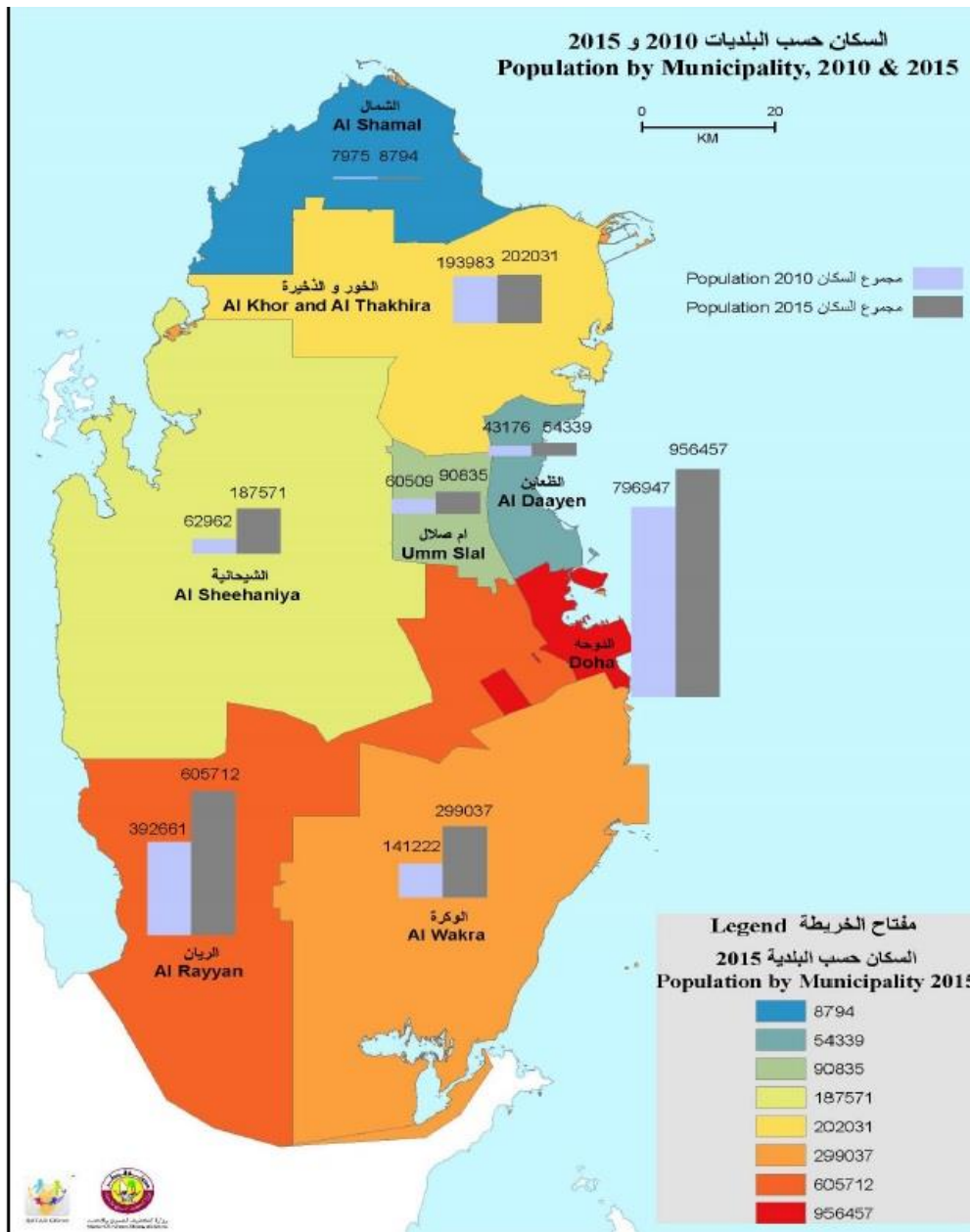


Figure 1.13. Map of State of Qatar showing the distribution of total population by municipality between 2010 and 2015. [Taken from Ministry of Development Planning and Statistics (2015)].

The increase in population is expected to continue to rise until 2020. When the main construction projects will be completed, after which the number will decrease gradually. The capital investment in mega infrastructure projects, which are funded with Qatar’s hydrocarbon revenues, including the Rail project, Hamad Port, Hamad International Airport, FIFA 2022 stadiums and facilities associated with this global event, and schools and hospitals, which have led to economic growth (Planning and Statistics Authority, 2019b).

In 2018, the labour force in Qatar amounted to 2.097 million economically active people. The results showed a rise of 2% compared to the previous year which was 2.057 million. The non-Qatari labour force annual growth rate increased by 2% and they represented 95% of total labour force in 2018. Around 79% of labour force work in the private sector, 10% work in the government sector, whereas the remaining work in mixed and domestic sectors (see Figure 1.13 and Figure 1.14).

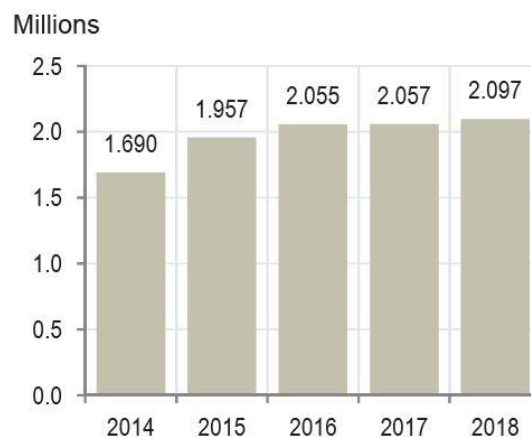


Figure 1.14. Labour force (15 years and above), in millions (2014-2018) [Taken from Planning and Statistics Authority, 2019a].

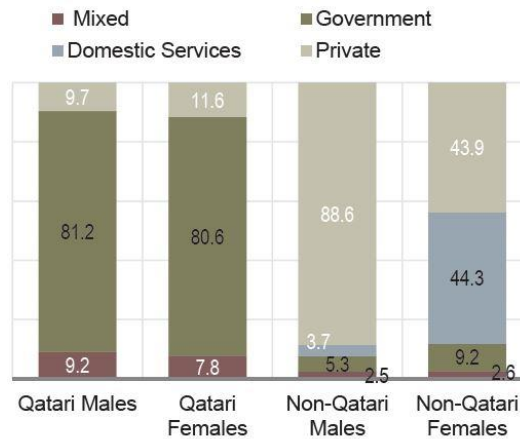


Figure 1.15. Bar charts showing the relative distribution of labour Force by nationality, sex and sector, 2018 [Taken from (Planning and Statistics Authority, 2019a)].

1.7.2 Forensic casework in Qatar

Forensic Biology and DNA section of the Forensic Laboratory Department is responsible for examining recovered evidence for the presence and identification of body fluid stains, perform DNA profiling of casework samples and reference samples, and paternity testing.

In 2011 DNA STR-based Database Unit was initiated, and the year 2013 witnessed the issuance of the Qatar DNA Law No (9). Currently, in Qatar STR typing is the main tool for forensic DNA analysis using a powerful commercial kit, GlobalFiler™ (TFS). Also, for particular cases such as paternity, male-lineage and sexual cases YFiler™ Plus Kit (TFS) is used. However, some highly degraded samples fail to produce a full profile or sometimes there is no profile. Single nucleotide polymorphisms (SNPs) typing can be considered as a supplementary tool with the routine STR workflow for analysing difficult samples, especially when high levels of degradation are present.

In summary, this research will study SNP typing using the Ion Torrent Personal Genome Machine (PGM) (TFS). The study will evaluate the MPS panels which are available for forensic applications: Precision ID Identity panel and Precision ID Ancestry.

Working hypothesis: MPS technology can enhance the capacity of forensic genetics in Qatar by providing additional markers that have been selected for either identity or ancestry. The identity SNPs can increase the discriminatory capacity of DNA profiling for individual identification and the ancestry SNPs can provide information on the biogeographical origins of biological material that have not be matched to an individual.

1.8 Project Aims

This project was designed mainly to evaluate the massively parallel sequencing within the Forensic Laboratory of Qatar through using ready-to-use MPS panels, the Precision ID Ancestry Panel and the Precision ID Identity Panel through Ion Torrent PGM sequencer. The rationale was to assess the MPS workflow and their potential use as an identification tool in conjunction with the routine STR workflow in Qatar Forensic Lab.

The specific objectives were:

- a) To evaluate the sensitivity of both MPS panels the Precision ID Ancestry panel and the Precision ID Identity panel. For this experiment 5 different DNA concentrations were used: 1 ng, 0.5 ng, 0.25 ng, 0.05 ng and 0.01 ng.
- b) To evaluate the massively parallel sequencing workflow using Ion Torrent PGM with both panels for processing reference samples and casework samples.
- c) To genotype 300 unrelated Qatari individuals from the eight municipalities that makeup the State of Qatar using the Precision ID Ancestry panel.
- d) To investigate ancestry inference ability of the Precision ID Ancestry Panel in Qatar population and to evaluate whether the Precision ID Ancestry Panel is a useful tool for ancestry inference in criminal casework.
- e) To genotype 105 unrelated Qatari individuals collected from the eight municipalities using Precision ID Identity panel.
- f) To evaluate the Precision ID Identity panel performance on aspects of heterozygote balance (Hb) and noise level (NL) and coverage values.
- g) To determine Precision ID Ancestry Panel and the Precision ID Identity Panel success in typing forensic casework samples. A total of 76 cases, which include 148 real casework samples were collected from different security departments dating from 2005-2018.
- h) To evaluate the forensic statistical parameters generated from the 300 Qatari individuals and the forensic parameters for the 165 ancestry SNPs.
- i) To evaluate the population structure using STRUCTURE for the eight municipalities studied.

- j) To evaluate the forensic statistical parameters based on 90 autosomal SNPs included in Precision ID Identity panel generated from the 105 Qatari individuals and their utility in casework analysis along with STR standard analysis marker in Qatar forensic DNA lab.
- k) To investigate the Y-haplogroup distribution among Qatar population. The 34 Y-SNPs included in the Precision ID Identity Panel were used to determine Y-haplogroups among 84 male samples from a total 105 Qatari samples that were used in Precision ID Identity Panel experiment.
- l) To combine the results from the two MPS panels applied to the casework samples and to evaluate the generated data to assess its importance in routine casework analysis.
- m) To evaluate Ion chef system performance for template preparation, automation library preparation and chip loading.
- n) To analyse the data and write up the PhD thesis

CHAPTER: 2 Materials and Methods

2.1 Overview

This chapter presents the materials and methods used in all the studies and experiments performed in this thesis.

The risk of contamination was minimized by wearing personal protective equipment (PPE) during all the routine work including lab coats, disposable gloves, and face masks. The bench surface and pipettes were cleaned with 10% bleach (sodium hypochlorite (v/v) and 70% ethanol. All the extraction steps were done with a negative extraction control. Negative control and positive controls were included in all the PCR reactions. The reagents were aliquoted in 30 ml plastic tubes (Fisher Scientific) and all the consumables were used exposed to UV light in a UV Crosslinker (Table 2.1).

Table 2.1 List of frequently used chemicals and reagents.

Item	Supplier
Bode SecurSwab DNA Collection System	Bode Technology
Phenol : Chloroform: Iso-amyl alcohol, (25 : 24 : 1)	ThermoFisher
PrepFiler Express™ Forensic DNA Extraction kit	
PrepFiler Express BTA™ Forensic DNA Extraction Kit	
Quantifiler™ Trio DNA Quantification Kit	
MicroAmp®96 Well reaction plate	
MicroAmp™ Clear Adhesive Film	
MicroAmp™ Optical Film Compression Pad	
Precision ID Ancestry Panel	
Precision ID Identity Panel	
Precision ID Library Kit	
Precision ID DL8 Kit	
Ion PGM™ Hi-Q™ View Chef Kit	
Ion 316™ Chip v2 BC	
Ion Xpress™ Barcode Adapters 1–16 Kit, Ion Xpress™	
Barcode Adapters 17–32 Kit	
Ion PGM™ Hi-Q™ View Sequencing Kit	
Wash 2 Bottle kit	
Ion Library TaqMan™ Quantitation Kit	
Agencourt™ AMPure™ XP Kit	Beckman Coulter™
QIAamp DNA Micro Kit	Qiagen
QIAamp® DNA Blood Maxi Kit	
EZ1 DNA Investigator Kit	
MaXtract High Density	
Chelex® 100	Sigma- Aldrich
Microcons	Merck Millipore

2.2 Ethics approval and permission

The study was approved by STEMH ethics committee of the University of Central Lancashire (Appendix 2). All the experimental work was done in the Forensic DNA and Biology Section in the Forensic Laboratory Department, Doha with a permission letter.

2.3 Sample collection

2.3.1 Qatari population samples

2.3.1.1 The strategy

Three hundred DNA samples were obtained from Qatari individuals living in the 8 municipalities that make up the State of Qatar.

- Different Departments in the Ministry of Interior were visited and the majority of the Ministry's employees are Qataris living in different parts of the country. In addition, a visit was undertaken to the Police College where the two cohorts of its students were Qataris from different areas of Qatar (Table 2.2).
- An official letter was sent to each department before each visit.
- A small group of participants that are not employees of the Ministry also participated in this study.
- After the completion of the sample collection a visit to The Centre for Geographic Information Systems (GIS) in Ministry of Municipality and Environment was undertaken to get official maps of the State of Qatar with all zones and municipalities (See Appendix 1 for Qatar Map).

Table 2.2. Data showing the total number of samples collected from the Ministry of Interior departments and from other volunteers.

Department Names	The number of samples
Police College	173 samples
Criminal Evidences and Information Department	24 samples
Human Resources Department	21 samples
Medical Services Department	14 samples
Forensic Laboratory Department	42 samples
Other	26 samples
Total	300

2.3.1.2 The samples

Buccal swabs from 300 Qatari volunteers (253 male and 47 female) were collected: all individuals self-identified as Qatari. The samples were collected from Qatari individuals from the eight municipalities (Table 2.3). The participants were given a Participant Information Sheet, which explained the work and why they had been asked to provide a sample, and completed an Informed Consent Form. Firstly, each sample was given a unique barcode (4 copies) and the sample was labelled with one barcode on the buccal swab. Secondly, one barcode was placed on the consent form. Thirdly, another barcode was used placed on the extraction sheet and fourthly, a barcode was given to volunteer. Should the volunteer decided to withdraw from the study they could have cited the barcode to identify their swab. All the samples were labelled as the following (from QAT1 to QAT300).

Table 2.3. Data showing the total number of samples collected from different municipalities; each municipality was coloured according to its colour on the map attached in Appendix 1.

Names of Municipality	The number of samples
Doha Municipality	75 samples
Al Rayyan Municipality	132 samples
Umm Salal Municipality	37 samples
Al Khor and Al Thakira Municipality	12 samples
Al Sheehaniya Municipality	10 samples
Al Wakra Municipality	20 samples
Al Shamal Municipality	5 samples
Al Daayen Municipality	9 samples

2.3.2 Forensic casework samples

2.3.2.1 The strategy

The aim was to collect real non-probative casework samples from different cases and geographical areas in the State. As samples are sent to the Forensic Laboratory Department from different security departments that are located in different geographical area and zones in the state, each security department was considered as a geographical area as follows (Table 2.4):

- Capital Security Department (as Doha Municipality).
- Rayyan Security Department (as Alrayyan Municipality)
- North Security Department (as Alshamal, Al Khor and Al Thakira, Umm Salal, and Al Daayen Municipality).
- Dukhan Security Department (as Al Sheehaniya Municipality).
- South Security Department (as Alwakra Municipality).

Table 2.4. Data showing the total number of casework samples collected from different municipalities.

Municipality Name	The number of samples
Doha Municipality	55
Al Rayyan Municipality	51
Umm Salal Municipality	4
Al Khor and Al Thakira Municipality	9
Al Sheehaniya Municipality	1
Al Wakra Municipality	25
Al Daayen Municipality	2

2.3.2.2 The samples

A total of 76 cases, which included 148 real casework samples were collected from different security departments from 2005-2018. The samples comprised different biological evidence types such as bloodstains, semen stains, hairs, saliva stains, teeth, bones, tissues and touch evidence (Table 2.5).

Table 2.5. Data showing the total number of different biological samples collected.

Sample Types	The number of samples
Hairs	4
Cells	33
Semen	6
Bones	6
Teeth	2
Blood	33
Saliva	26
Touch	26
Tissues	12
Total	148

2.3.2.2.1 Sample preparation

Before to performing DNA profiling test, all samples were examined with several initial tests to determine the type of biological material. For bloodstains, Kastel- Mayer test was used to detect blood; the possible presence of haemoglobin. Also, the ABACard®Hematrace® (Abacus Diagnostics) test, which was used to aid in the possible identification of human blood by detecting the presence of human haemoglobin. Phadebas® Amylase (Maggie Life Sciences) test was applied on saliva stains to indicate the presence of saliva.

Several tests were applied on the samples that contained semen stains. Firstly, Ultraviolet (UV) lamp was used for locating possible dried semen stains. Secondly, the Acid Phosphatase test using Phosphatesmo KM (Macherey-Nagel) was used for the presumptive identification of semen in stains through the presence of acid phosphatase enzyme. Next, ABACard P30 (Abacus Diagnostics) is a confirmatory test and was performed to detect the human prostate-specific p30 protein found in seminal fluid. The final step to examine samples that contained semen stains was the microscopic examination of slides for the visualisation of spermatozoa and nucleated epithelial cells using Haematoxylin and Eosin (H&E) stain.

For hair, the samples were prepared for microscopic examination. Each hair was mounted with sterile water and a coverslip to examine whether the hair is human hair or not and also to test the suitability of a human hair sample for nuclear DNA analysis. Bones and teeth samples were selected for the study were prepared through several cleaning steps before the DNA extraction protocol started. The specimens were cut into small fragments by using a cleaned cutting wheel fitted to Dremel rotary tool. The outer and the exposed layer of the sample was cleaned with a sanding paper, Dremel sanding bits were used for some samples. The sanded samples were placed into 50 ml conical tubes and washed with 10% bleach solution (Sodium Hypochlorite) for 3 s 3 times. The bleach was decanted and the samples were washed with 70 % ethanol for 3 s 3 times. The 70% ethanol was decanted and the cleaned fragments were placed in a sterile petri dish for a drying step in 56 °C oven for approximately 2 h. The powdering step was carried out using the TissueLyser machine (Qiagen). Following the TissueLyser operating guide, the samples were placed into cleaned steel grinding sets. Then the powdered samples were transferred to measuring pots to collect the amount needed for the extraction and any extra sample powder was kept at -20 °C for any future processing.

2.4 Extraction

The 300 population samples were extracted using PrepFiler Express™ chemistry (TFS) on an Automate Express DNA extraction system (TFS). The selected casework samples were isolated by different routine casework extraction methods, manual and automation protocols used in the Forensic Biology and DNA sections. Manual extraction was by either organic Phenol-chloroform-isoamyl alcohol, Chelex 5%, or QIAamp DNA Micro Kit (Qiagen). The automation extraction systems with EZ1 DNA Investigator Kit (Qiagen), PrepFiler Express™ or PrepFiler Express BTA™ Forensic DNA Extraction Kits (TFS) (Table 2.6).

Table 2.6. Table showing total number of samples according to their extraction methods.

Extraction methods	The number of samples
Phenol-chloroform-isoamyl alcohol	74 samples
QIAamp Mini columns	38 samples
PrepFiler Express	16 samples
PrepFiler Express BTA™ kit	8 samples
Chelex 5%	7 samples
EZ1 DNA Extraction	3 samples
Qiagen Maxi Kit	2 samples
Total	148 samples

2.4.1 Extraction of population samples

DNA was extracted from 300 buccal swabs using PrepFiler Express™ chemistry (TFS) on an Automate Express DNA extraction system. The substrate (half swab) was carefully transferred to a labelled PrepFiler LySep™ Column and 500 µl PrepFiler™ Lysis Buffer and 5 µl freshly-prepared 1 M DTT was added and the tube tightly closed. Then the samples were placed in the Thermomixer incubator at 70 °C and 750 rpm for 40 min. The column tubes were centrifuged for 2 min at 10,000 x g to transfer the lysate to the sample tube. The AutoMate Express™ Instrument was turned on to run the samples and onscreen instructions were followed. The elution tubes containing 50 µl extracted DNA was kept at 4 °C until use.

2.4.2 Extraction of casework samples

2.4.2.1 EZ1 DNA Investigator Kit

The Biorobot EZ1 instrument was turned on and following the instructions appeared on the screen, the UV decontamination protocol was selected to start the decontamination process. The samples (blood stains) were placed in 1.5 ml micro-centrifuge tube. A mix consisting of 190 µl G2 buffer (Qiagen) and 10 µl Proteinase K was added to the samples. Then the samples were vortexed and placed in the Thermomixer incubator at 56 °C and 900 rpm for 2 h. To collect the cells the samples were placed in spin basket (Promega Corporation) and were centrifuged for 5 min at 10,000 x g. The lysate was transferred to 2 ml EZ1 sample tubes and 1 µl Carrier RNA

(Qiagen) was added to each sample. Trace protocol was selected to extract the DNA. The DNA was eluted in 50 µl TE Buffer and was stored at 4 °C.

2.4.2.2 The PrepFiler Express™ and PrepFiler Express BTA™ Forensic DNA

Extraction Kits Kit

The samples (touch, cells, saliva and blood stains) which were selected for the study and extracted with PrepFiler Express™ Kit were extracted as the population samples explained above in Section 2.4.1.

2.4.2.2.1 The PrepFiler Express BTA™ Forensic DNA Extraction Kit

The kit was used to extract the DNA from chewing gum and cigarette butts, and bone samples.

Following the user guide, the adhesive substrates protocol was applied to extract the DNA from chewing gum and cigarette butt. The chewing gum was cut into pieces of approximately 5 mm in thickness and placed into a clean petri dish and kept at -80 °C for at least 2 h. 50 mg of the chewing gum was inserted into a labelled PrepFiler LySep™ Column. With cigarettes the outer paper surrounding the cigarette filter was cut into pieces and transferred into PrepFiler LySep™ Column. Whilst the protocol of bone and teeth was applied to extract the bone and teeth samples. Approximately 50 mg of bone and teeth powder were weighed and transferred into PrepFiler™ bone and tooth Lysate Tube. A fresh lysis mix was prepared and calculated as each sample required 220 µl PrepFiler BTA™ Lysis Buffer, 5 µl freshly-prepared 1 M DTT and 7 µl Proteinase K. To each sample, 230 µl of freshly prepared BTA solution was added. The tubes including chewing gums and cigarettes were incubated in the Thermomixer incubator at 56 °C and 750 rpm for 40 min. The tubes with bones were vortexed and centrifuged shortly then were placed in the incubator at 56 °C and 1100 rpm overnight. The tube containing the bone was centrifuged at 10,000 $\times g$ for 90 s and the clear supernatant was transferred to a new labelled PrepFiler™ Sample tube whereas the chewing gum and cigarette PrepFiler LySep™ Column were centrifuged at 10,000 $\times g$ for 2 min. The AutoMate Express™ Instrument was turned on and the PF Express BTA option was selected to run the samples and onscreen instructions were followed. The elution tubes containing 50 µl extracted DNA was kept at 4 °C until use.

2.4.2.3 Chelex extraction

The samples contained semen stains extracted with Chelex which were used in this study were extracted using the differential extraction method using 5% and 20% Chelex®100 (Sigma-Aldrich). The cut stains were placed in a new labelled 1.5 ml microcentrifuge tube. To each tube 1 ml of Phosphate-buffered saline (PBS) (8 g NaCl, 0.2 g KCl, 1.44 g Na₂HPO₄, 0.24 g KH₂PO₄) was added and vortexed for 10 s. The samples were incubated for 30 min at room temperature. The substrates were placed in a spin basket (Promega Corporation) which was fitted on the same micro-centrifuge tube and centrifuged at 10,000 *x g* for 5 min. Around 950 µl of the supernatant was discarded and only 50 µl of the supernatant was kept to cover the pellet. The pellet was re-suspended in 150 µl of sterile distilled water and 2 µl of Proteinase K then vortexed for 10 s and incubated at 56 °C for 1 h. The tubes were centrifuged at maximum speed 20,000 *x g* for 5 min. From the supernatant, 150 µl was transferred to a new 1.5 ml microcentrifuge tube and labelled as epithelial fraction (E1) and 50 µl of 20% Chelex®100 (Sigma-Aldrich) was added. The sperm pellet was washed with 500 µl of sperm wash buffer (SWB) and was vortexed and centrifuged at 20,000 *x g* for 5 min. The supernatant was removed and the pellet was washed with 450 µl of SWB for a second time. The sperm pellet was washed with 1 ml of sterile distilled water then the tube was vortexed and centrifuged at 20,000 *x g* for 5 min. Without disturbing the pellet, the supernatant was removed and the pellet re-suspended with 150 µl of 5% chelex®100 (Sigma-Aldrich) and 7 µl of 1.0 M DTT and 2 µl of Proteinase K. The tubes were vortexed and incubated at 56 °C for 1 h. The female (E1) and sperm fraction (E2) were placed in the Thermomixer incubator at 99 °C for 9 min. The tubes were centrifuged at maximum speed (20,000 *x g*) for 3 min. The supernatant was transferred to a new Microcon filter 100 KDa (Merck Millipore) fitted on a 0.5 ml tube. The tubes were centrifuged at (500 *x g*) for 15 min. A second wash was carried out with 200 TE buffer (10 mM Tris-HCl and 0.1 M EDTA pH 8) at 500 *x g* for 15 min. The filter was placed into a new labelled tube and 20 µl of TE buffer was added to the filter and incubated at room temperature for 5 min. The filter was flipped and centrifuged at 1020 *x g* for 5 min. The eluted DNA was stored at 4 °C until use.

2.4.2.4 Phenol-chloroform-isoamyl alcohol

The samples (tissue, touch, cells, hair root, saliva and blood stain) were cut and placed into a new labelled micro-centrifuge tube. A volume of 500 µl of Stain Extraction Buffer (SEB) (1 M Tris, 1M NaCl, 0.5 M EDTA and 20% sodium dodecyl sulphate) and 20 µl of Proteinase K were added. The samples were incubated at 56 °C for 2 h and for trace samples the incubation was overnight. The sample were transferred to a basket (Promega Corporation) which was fitted on the same 1.5 ml micro-centrifuge tube. The tube was centrifuged at 10,000 x *g* for 5 min. After the centrifugation the basket was removed and the supernatant was transferred to a new MaXtract High density gel tube (Qiagen). A volume of 500 µl of phenol-chloroform-isoamyl alcohol (PCI) was added to the sample and vortexed for 3 min. The tube was centrifuged at 16,000 x *g* for 5 min. The aqueous layer was transferred to a new Microcon filter fitted on a new micro-centrifuge tube. The tube was centrifuged at 500 x *g* or 15 min. A volume of 200 µl TE buffer (10 mM Tris-HCl and 0.1 M EDTA) was added to the filter and the tube was centrifuged at 500 x *g* for 15 min. The Microcon filter was placed onto a new microcentrifuge tube and a 20 µl of TE buffer (10 mM Tris-HCl and 0.1 M EDTA) was added. The tube was incubated at room temperature for 5 min and the filter was flipped and centrifuged at 1020 x *g* for 5 min. The DNA was kept at 4 °C until use.

2.4.2.5 Qiagen MicroAmp Kits

According to the QIAamp® DNA Investigator Handbook, the samples (tissue, touch, cells, hair root, saliva and blood stain) were cut and placed into a 2 ml micro-centrifuge tube. The sample were digested with 400 µl of ATL and 20 µl of Proteinase K and the tube was vortexed for 10 s. The tube was placed in a Themomixer and incubated at 56 °C for 2 h and an overnight incubation for the trace samples with shaking at 900 rpm. The swab heads/cut material was transferred into a new spin basket (Promega Corporation) fitted on the sample tube, and then centrifuged at 10,000 x *g* for 5 min. The basket was discarded and 400 µl AL and 1 µl of dissolved carrier RNA were added. The tube was plus vortexed for 15 s and incubated at 70 °C for 10 min with shaking at 900 rpm. The tube was briefly centrifuged and 200 µl of ethanol (96%) was added. Then, the tube was vortexed for 15 s and briefly centrifuged to collect the droplets.

Without wetting the rim, the lysate was transferred to a new labelled QIAamp MinElute Column, and was centrifuged at $6000 \times g$ for 1 min.

The QIAamp MinElute Column was placed onto a new 2 ml collection tube, whilst the collection tube containing the flow-through was discarded. The remaining lysate was transferred to QIAamp MinElute Column and centrifuged at $6000 \times g$ for 1 min. The QIAamp MinElute was placed onto a new 2 ml collection tube and the tube containing the flow-through was discarded. 500 μ l of AW1 was added to the QIAamp MinElute Column and the tube was centrifuged at ($6000 \times g$) for 1 min. The column was placed onto a new 2 ml collection tube and the old collection tube was removed and 500 μ l of AW2 was added. The tube was centrifuged at $6000 \times g$ for 1 min followed with another centrifugation step at maximum speed $20,000 \times g$ for 3 min to ensure that the column membrane was completely dry. The QIAamp MinElute was placed onto a new 1.5 microcentrifuge tube and 25 μ l of Buffer AE was added to the centre of the column's membrane. The column was incubated at room temperature for 1 min and centrifuged at higher speed $20,000 \times g$ for 1 min. The eluted DNA was kept at 4 °C.

2.4.2.6 QiaMaxi DNA Extraction

The QiaMaxi protocol was used to extract DNA from bones and teeth samples. Between 0.5 g to 3 g of powdered bone was placed in a 50 ml polypropylene tube. A volume of 20 ml of 0.5 M EDTA was added and the tube was vortexed and incubated overnight at 4 °C. The tube was centrifuged at $2000 \times g$ for 15 min, and the EDTA solution was decanted. The decalcification step was repeated till the sample was digested. The pellet was washed with 40 ml of sterile distilled water and the tube was vortexed and centrifuged at $2000 \times g$ for 15 min. The water wash was repeated for a total of three times and the water was decanted. A volume between 5 – 10 ml of ATL Buffer (Qiagen) and 150 – 300 μ l of Proteinase K (20 mg/ml) was added to each sample. The tube was briefly vortexed and the mixture was incubated at 56 °C overnight. The following day the tube was briefly centrifuged at $2000 \times g$ for 1 min. A volume between 5 – 10 ml of AL Buffer (Qiagen) and 150 – 300 μ l of Proteinase K (20mg/ml) was added to each sample lysate. Then, the tube was vortexed vigorously for 10-15 s, followed by a short pulse spin to remove the detergent from the top of the

cap and the sample was incubated at 70 °C for 1 h. The mixture was centrifuged at 2000 x *g* for 5 min, then the supernatant was transferred to a new labelled 50 ml conical tube filled with a volume of 5-10 ml of 100% ethanol. The tube was vortexed for 10-15 s. The entire mixture was transferred to a clean, labelled QiaMaxi column was centrifuged for 3 min at 1850 x *g*. The collection tube was discarded and the QiaMaxi filter was placed into a new waste collection tube. Any remaining mixture was added to the column which was then centrifuged at 1850 x *g* for 3 min. The collection tube was discarded and the QiaMaxi filter was fitted onto a new collection tube. 5 ml of AW1 buffer was added to the QiaMaxi filter and the tube was centrifuged at 4500 x *g* for 1 min. The filtrate was discarded, and the column was placed in a clean collection tube and 5 ml of AW2 buffer was added into the column. The tube was centrifuged again at 4500 x *g* for 20 min. The waste collection tube was discarded and the QiaMaxi filter was fitted onto a new 50 ml conical collection tube. A volume between 0.5-1 ml of TE was added to the column membrane to elute the DNA. The column was incubated at room temperature for 5 min and then centrifuged for 2 min at maximum speed 4500 x *g*. An addition volume of 1 ml of TE was added to QiaMaxi filter membrane and then incubated at room temperature for 5 min. The tube was centrifuged at maximum speed 4500 x *g* for 5 min. The eluted DNA was transferred to a new labelled Vivacon® filter 100K device (Viva-products) and was centrifuged at 2500 x *g*. The amount of solution was checked every 4 min until most of the solution had flowed through. The reservoir was inverted and collection tube portion into the cylindrical-shaped tube containing the filtrate and then was centrifuge for 2 min at 3000 x *g* to collect the DNA. The DNA extract was transferred to a new labelled 1.5ml micro-centrifuge tube and kept at 4 °C.

2.5 DNA Quantitation

The extracted DNA samples were quantified using Quantifiler® Trio DNA Quantification Kit (TFS) on a 7500 Real Time PCR and following the Quantifiler® HP and Trio DNA Quantification Kits user guide. The forensic casework samples that were selected for the study and were not originally quantified with Quantifiler® Trio DNA Quantification Kit, were re-quantified with Quantifiler® Trio DNA Quantification Kit. The quantification was performed using 18 µl of mastermix, 8 µl of Quantifiler® Trio Primer

Mix and 10 µl of Quantifiler® THP PCR Reaction Mix. Five 0.2 µl tubes were labelled with Std1, Std2, Std3, Std4 and Std5 to prepare a ten-fold dilution series ranging from 50.0 ng/µl to 0.005 ng/µl. Thereafter, 2 µl of DNA/Standard and for the NTC 2 µl of Quantifiler® THP DNA Dilution Buffer were added. The plate was sealed and placed in the centrifuge for 20 s at 1811 x *g* then place in the 7500 Real-Time PCR instrument.

2.6 Library preparation and Quantitation

Library was prepared in this study manually and automatically using the Ion Chef instrument using the Precision ID Library Kit. The targets were amplified using two panel Precision ID Ancestry Panel and Precision ID Identity Panel (Table 2.7).

Table 2.7. Data showing the total samples analysed with Precision ID Ancestry Panel and Precision ID Identity Panel.

Panels	Population samples	Casework samples
Ancestry Panel	300 samples	148 samples
Identity Panel	105	60

2.6.1 Manual library preparation and quantitation

2.6.1.1 Amplification the targets

In total, 576 samples were prepared manually. The Precision ID Ancestry Panel was applied on 300 Qatari reference population samples and 111 casework samples. From the 300 Qatari samples, 105 Qatari samples were selected from the different area and zones from the state and amplified with Precision ID Identity Panel (Table 2.8). Also, the Precision ID Identity Panel was applied on 60 casework samples.

Table 2.8. Data showing the 105 Qatari samples studied with Precision ID Identity Panel.

Names of the Municipalities	The number of samples
Doha Municipality	20 samples
Al Rayyan Municipality	25 samples
Umm Salal Municipality	15 samples
Al Khor and Al Thakira Municipality	11 samples
Al Sheehaniya Municipality	10 samples
Al Wakra Municipality	11 samples
Al Shamal Municipality	5 samples
Al Daayen Municipality	8 samples

Following the Precision ID Panels with Ion PGM System application guide supplied by ThermoFisher Scientific, Precision ID Library Kit (96Rxn) was used to build the libraries. The samples were amplified with an input of 1 ng DNA using Precision ID Ancestry Panel and Precision ID Identity Panel. A PCR mastermix was prepared which consisted of 10 μ l of Precision ID Ancestry Panel and Precision ID Identity Panel with 4 μ l of 5X Ion Ampliseq HiFi Mix per sample and 14 μ l of the mix was aliquoted into each well of the 96 well PCR plate. All the samples were adjusted to a concentration of 1 ng/ μ l and 1 μ l was added to each well and 5 μ l of nuclease-free Water. This was applied for the population samples and forensic casework samples when their DNA concentration was 1 ng/ μ l or above. Whereas casework samples which were below 1 ng/ μ l, were amplified with a full volume of 6 μ l DNA. The plates were covered with MicroAmp Adhesive Film followed by a vortex step and short spin in the centrifuge to collect any droplets. The ABI thermal cycler 9700 was programmed using the conditions outlined in Table 2.9 and Table 2.10. The number of PCR cycles was adjusted according to the DNA input. Following the application guide 21 cycles were for the samples with 1 ng and the PCR cycles number was increased and 26 cycles for the casework samples < 1 ng.

Table 2.9. Data showing the PCR run programme used for the samples with 1 ng.

Stages	Steps	Temperatures	Time points
HOLD	Enzyme Activation	99 °C	2 min
21 Cycles	Denaturing	99 °C	15 s
	Annealing/ Extending	60 °C	4 min
Hold	-	10 °C	Hold

Table 2.10. Data showing the PCR run programme used for the samples (casework samples) with <1 ng.

Stages	Steps	Temperatures	Time points
HOLD	Enzyme Activation	99 °C	2 min
26 Cycles	Denaturing	99 °C	15 s
	Annealing/ Extending	60 °C	4 min
Hold	-	10 °C	Hold

2.6.1.2 Primer partial digestion

The resulting amplified samples were treated with 2 µl of FuPa reagent one of the library component kit. FuPa reagent was added to partially digest the primers and phosphorylate the amplicons for the ligation step. The plate was sealed with an Optical Adhesive Cover and placed in ABI thermal cycler 9700 and programmed as shown in Table 2.11. The final volume was 22 µl.

Table 2.11. Data showing partial digestion PCR programme.

Temperatures	Time points
50 °C	10 min
55 °C	10 s
60 °C	20 min
10 °C	Hold for up 1 h

2.6.1.3 Ligation of Barcode Adapters to the digested samples

All the digested amplified products were ligated with Ion Xpress™ P1 Adapter and Ion Xpress™ Barcode Adapters (1-16 kit) and (17-32 Kit), each barcode consisted of 8-10 nucleotides and the P1 adapter. According to the manufacture's user guide a dilution of 1:4 needed to be prepared for each barcode adapter as detailed in Table 2.12. The mix of the diluted barcode adapters could be stored at -20 °C.

Table 2.12. Data showing an example of diluted barcode adapter mix for 4 runs.

Component	Volume
Ion P1 Adapter	2 µl
Ion Xpress Barcode X	2 µl
Nuclease- free Water	4 µl
Total	8 µl

The ligation step was done by adding 2 µl of diluted barcode adapter mix, 4 µl of Switch Solution and 2 µl DNA Ligase. The plate containing the amplicons was sealed, vortexed and pulse spun and placed in the thermal cycler with the programme shown in Table 2.13.

Table 2.13. Data showing the thermal cycler programme for ligation step.

Temperature	Time
22 °C	30 min
72 °C	10 min
10 °C	Hold for up to 1 h

2.6.1.4 Purification of unamplified libraries

1. Following the Precision ID Panels with Ion PGM System application guide; 70% ethanol was freshly made for this step by mixing 230 µl of 96% ethanol and 100 µl of nuclease-free water for each sample.
2. The plate containing barcoded libraries was purified by adding 45 µl of Agencourt® AMPure® to each well. The plate was sealed by MicroAmp Clear Adhesion Film, vortexed and briefly centrifuged to collect any droplets. The plate was incubated at room temperature for 5 min.
3. The plate was placed in a DynaMag™-96 Side Magnet a magnetic rack and incubated at room temperature for 2 min. The clear supernatant was carefully removed.
4. A volume of 150 µl of freshly prepared 70% ethanol was dispensed to each well containing the libraries, the beads were washed by moving the plate from one side to another and the cleared supernatant was discarded. The ethanol washing step was repeated once.
5. The plate was placed in the centrifuge for a short spin and placed in the magnetic rack until the well was clear. The remaining ethanol drops were removed by using a pipette.
6. The plate was placed in the magnetic rack and dried in air at room temperature.
7. The plate was removed after the drying and 50 µl of Low TE was added to the pellet to disperse the beads.
8. The plate was sealed with MicroAmp Adhesive Film pulse vortexed and centrifuged for 5 min at 657 x g. The samples with beads were kept at 4 °C for up

to one month while for long storage the beads were removed and the purified unamplified libraries were transferred to a new plate and stored at -20 °C.

2.6.1.5 qPCR quantitation

The concentration the unamplified libraries was determined and quantified by qPCR. A dilution 1:100 was prepared for the samples by adding 2 µl of purified sample with 198 µl of Nuclease-free Water. *E. coli* DH10B Ion Control Library was used to prepare three standards of 6.8 pmol (Std 1), 0.68 pmol (Std 2) and 0.068 pmol (Std 3). The master mix was prepared of 10 µl of Ion Library TaqMan® qPCR Mix and 1 µl of Ion Library TaqMan® Quantitation Assay, 20x and 11 µl was aliquoted for each well. 9 µl from the standards and the diluted libraries was transferred to each well. NTC well was filled with 9 µl Nuclease-free Water. The quant plate was sealed and placed in 7500 thermal cycler system and the run was set according to the manufacture's user guide.

2.6.1.6 Template dilution factor (TDF) determination

The real-time PCR data results were analysed and qPCR quantity mean was used to calculate the average concentration of the undiluted library according to the manufacture's user guide formula (qPCR quantity mean) X (library dilution (100)), to calculate the template dilution factor (TDF) which was needed to prepare a final concentration of 50 pM or 30 pM. The first batch of samples studied, which were the 300 population samples, were calculated at 50 pM. The second batch of the samples, 148 casework samples and the 105 Qatari samples were calculated at 30 pM (based on communication from ThermoFisher Scientific). All the libraries samples were pooled in equimolar amounts and were loaded to the Ion Chef Sample tube.

2.7 Automation Library Preparation

A total of 48 casework samples were used to prepare DNA libraries with Ion Chef automation. Precision DL8 kit (Table 2.14) was used in library preparation automatically and with Precision ID Ancestry Panel (TFS).

Table 2.14. Data showing the Precision DL8 kit contents.

Item names	Storage temperatures
1 set of 4 IonCode in PCR plates (the set includes: IonCode 0101-0108 in 96-well PCR plate (red), IonCode 0109-0116 in 96-well PCR plate (yellow), IonCode 0117-0124 in 96-well PCR plate (green), IonCode 0125-0132 in 96-well PCR plate (blue))	Room temperature.
4 cartridges Ion AmpliSeq Chef reagents DL8	-5 °C to -30 °C
4 cartridges Ion AmpliSeq Chef solutions DL8	Shipped at room temperature and stored at 2–8 °C
4 boxes Ion AmpliSeq supplies (each box contains: Ion AmpliSeq tip cartridge L8, framed PCR foil seal, enrichment cartridge), store at room temperature	Room temperature.

2.7.1 Sample preparation

1. All the 48 casework samples were sequenced on two 316 chips. As each chip holds 24 samples, chip#1 was used for 24 samples with DNA concentration ≥ 1 ng while chip#2 was used 24 samples with DNA concentration < 1 ng.
2. The Precision ID DL8 IonCode™ Barcode Adapters Plate was used in automated DNA Library Preparation.
3. The plate seal was removed and discarded and in the first column (A1-H1) 15 μ l of DNA samples were added.
4. For samples which had low DNA concentration (below 1 ng) 15 μ l of each sample was pipetted in each well.

5. Precision ID DL8 Reagents cartridge tube 1 (position A) and tube 2 (position B) were opened and 150 µl of Precision ID Ancestry Panel primer was pipetted in each tube.

2.7.2 Creation of sample set

1. A sample set was created in Torrent Suite™ Software (TSS) by entering sample's details manually for a total of 8 samples per run in library preparation through Ion Chef.
2. The required sample information which was filled in TSS was: Sample name, Sample position in the IonCode 96-well PCR (A1, B1, C1, D1, E1, F1, G1, H1), Barcode kit and Barcode number as 0101 (sample in position A1 with a barcode 0101) and saved as a sample set with specific name.
3. Eight samples entered as a sample set were selected from the software list to proceed with the library preparation and provided the necessary instructions to Ion Chef regarding Sample Set Name, Library Prep Type, Library Prep Kit and PCR Plate Serial Number to start preparing libraries.

2.7.3 Running the Ion Chef

The ion chef instrument was turned on and the connection with software was checked. Following the Precision ID Panels with the Ion PGM™ System application guide and the instructions appeared on the Ion Chef touch screen, all the barcoded reagents, cartridges and consumable plastic materials were installed in Ion Chef in its allocated deck position, such as

1. The Precision ID DL8 Solutions cartridge into the Solutions station.
2. The Precision ID DL8 Reagents cartridge into the Reagents station.
3. A new Ion AmpliSeq™ Tip Cartridge DL8 into the New Pipette Tip station.
4. An empty pipette tip rack from a previous run into the used tip station.
5. Precision ID IonCode™ 96 Well PCR Plate containing the samples were loaded from A1-H1 wells and placed onto the thermal cycler block in the Ion Chef.
6. A new PCR foil seal was placed at the bottom of the automated heated cover.
7. Enrichment Cartridge into the Enrichment station.

The Ion chef door was closed to start the experiment and the saved sample set was chosen from the instrument touch screen. The run was performed by choosing the option step by step. For the automation library preparation, the option Ampliseq was chosen. A deck scan was performed before the run started the scanning process, the Ion Chef checked all the cartridges and the consumable plastic materials barcodes that had been entered in TSS in the planned sample set. The server information connection was verified and the name of the sample set appeared and was selected. According the user guide and the suitable number of primer pools, number of amplification cycles and the extension time was selected (Table 2.15). The tubes which contained DNA samples with 1 ng, the amplification cycle number was 22. Whereas for samples less than 1 ng the amplification cycle number was 25. After all the options been selected, the Ion chef run was started with a total 7 run time of h.

Table 2.15. Data showing the panel recommendation for the automated library preparation with Ampliseq workflow.

Panels	Amount of DNA		Cycle numbers	Anneal and Extension time
	Input gDNA added	# of Primer pools		
Precision ID SNP panels	1 ng (300 copies)	1	22 cycles	4
(Identity and Ancestry)	<1 ng (<300 copies)	1	22 cycles + 1 to 5 cycles	4

2.7.4 Unloading the Ion Chef and libraries dilution

After the Ion Chef Ampliseq run was complete the tube in position D (output tube) in the reagent station was removed, capped, labelled and was kept in -20 °C for the next run. Then, all the reagents and the consumables were removed and discarded from each station; only the empty Tip Cartridge was transferred to the Used Pipette Tip station. The output tube contained pooled barcoded libraries with a volume 700 µl at a concentration of approximately 100 pM. For Ion Chef-based templating the libraries were diluted to 30 pM.

2.8 Template preparation on the Ion Chef

2.8.1 Creation a planned run

The Precision ID Panels with the Ion PGM™ System application guide was followed for the planned run creation. The Torrent Suite browser page was opened and Template was selected. From the favourites menu Human Identification option was chosen. The Human Identification templates includes all the templates for different Applied Biosystems Precision ID Panels. The samples were run with Precision ID Ancestry Panel, the Applied Biosystems Precision ID Ancestry Panel – PGM Template was selected whereas the Applied Biosystems Precision ID Identity Panel – PGM Template was selected to those samples were prepared with Precision ID Identity Panel. For each run, the appropriate options were selected from the Kits list (Table 2.16). Also, for each panel the appropriate analysis parameters were selected (Table 2.17) and the required field were completed, including the Run Name, Number of Barcodes (total number of samples barcoded), Sample Tube Label (the tube containing the pooled libraries), Chip Barcode, the sample name for each barcode. After all the required fields were filled the plan run was saved and listed in the planned run page for the Ion Chef run.

Table 2.16. Data showing the appropriate information entered for each run.

Instrument	Ion PGM™ System
Chip Type	Ion 316™ Chip v2
Library Kit Type	Precision ID Library Kit
Template Kit	Ion PGM Hi-Q Chef Kit (Ion Chef selected)
Templating Size	200
Flows	500

Table 2.17. Data showing the analysis parameters for each panel.

Reference Libraries	Hg 19 (hg 19 from Zip)	
Target Regions	PrecisionID_IdentityPanel_targets.beds	Identity Panel
	PrecisionID_AncestryPanel_targets.beds	Ancestry Panel
Hotspot Regions	PrecisionID_IdentityPanel_hotspots.beds	Identity Panel
	PrecisionID_AncestryPanel_hotspots.beds	Ancestry Panel

2.8.2 Running the Ion Chef

Ion Chef was turned on and the connection to the TSS was checked. Each templating run was performed contained two 316 chips (Table 2.18). Following the user's guide and the instructions that appeared on the Ion chef touch screen, all the barcoded reagents, cartridges and consumable plastic materials were installed in each appropriate station in the Ion Chef such as Ion PGM™ Hi-Q™ Chef reagents cartridge and the tubes were opened. Ion PGM™ Hi-Q™ Chef solutions cartridge, a New Tip cartridge v2, an empty pipette tip rack were placed in the waste pipette tip (from previous run). A new PCR plate was placed into the Thermal cycler block and a new PCR Frame seal v2 was placed at the automated heated cover. The Enrichment cartridge v2 with six recovery tubes (v2) were placed into each recovery centrifuge and the two recovery centrifuges were closed with two recovery station lids disposable plastic covers. A two-chip adapter had been attached to the two Ion chips to avoid any loading failures. The chip adapter wells aligned to the chip's wells then the adapter pressed onto the chip to lock the chip till a small click was heard. The two chips were placed in the chip loading centrifuge bucket. The Ion chef door was closed. A deck scan was performed before the run started through the scanning process. The Ion Chef checked all the cartridges and the consumable plastic materials barcodes that had been entered in TSS in the planned run which was pending. The run started once the scan check was complete and successful. The planned run was chosen from the Ion Chef Touchscreen.

Table 2.18. Data showing the total number of samples templating on Ion Chef manually and automatically.

Sample types	Library Preparation	Total number of chips	Panels
300 Population samples	Manual	10 chips	Ancestry Panel
105 Population samples	Manual	4 chips	Identity Panel
111 Casework samples	Manual	5 Chips	Ancestry Panel
48 Casework samples	Automation	2 Chips	Ancestry Panel
60 Casework samples	Manual	2 Chips	Identity Panel

2.9 Cleaning the Ion PGM sequencer

The Ion PGM sequencer must be ready to sequence the chip at least one hour before the Ion Chef Machine run finished. Before starting the sequencing and loading the chips, two washing steps were done and the steps were performed by following Precision ID Panels with the Ion PGM™ System application guide. For the first sequencing run, a chlorite cleaning step and a water 18 MΩ step was done before each sequencing run followed by an initialization step. Before the washing step was started, the wash and the reagent bottles were removed from the Ion PGM™ sequencer. An old chip was placed in the chip socket and used as a cleaning chip. The 250 ml bottle Wash 1 and Wash 3 and the 2 l bottle Wash 2 were washed with 18 MΩ water 3 times then the bottles were labelled as washing bottles.

2.9.1 Condition the Wash 2 Bottle for First Use

For at least 8 h before the first use the new wash 2 bottles were conditioned with wash 2 bottle conditioning solution. Wash 2 was conditioned by filling the bottle with 18 M Ω water to the line present on the bottle. All the Wash 2 Bottle Conditioning Solution was added to the water. The bottle was closed and was mixed by inverting the bottle 5 times. The bottle was placed at room temperature for at least 8 h or overnight.

2.9.2 Chlorite Cleaning

In order to perform the washing step, a stock of 1 M NaOH was prepared weekly by mixing 10 M NaOH with 18 M Ω water. In addition, 100 mM of NaOH was prepared (on a daily basis) by mixing 1 M stock in 18 M Ω H₂O. Two 250 ml bottles and one 2 l bottle were labelled with Chlorite wash were used to perform this wash. The bottles were washed with 100 ml of 18 M Ω water 3 times. A glass bottle was filled with 1 L of 18 M Ω water and an Ion PGM™ cleaning tablet (Chlorite Tablet) was added. The glass bottle was kept for 10 min to dissolve the tablet. About 1 ml of 1 M NaOH was added and by using 0.45- μ m vacuum filter (Sigma-Aldrich) the solution was filtered. From the ION PGM touchscreen the clean option was selected and Chlorite Wash option was then selected. Following the instruction from the ION PGM touchscreen, an old chip (cleaning chip) was placed in the Chip Clamp. All wash and reagent bottles were removed from the PGM. The old sippers attached were kept in place at all positions. The 250 ml cleaning bottle was filled with 250 ml of filtered chlorite solution. The sipper attached to W1 position was cleaned with 18 M Ω water. The 250 ml bottle with the filtered chlorite solution was attached to the W1 position. The empty 2 l cleaning bottle was placed in the W2 position whereas the empty 250 ml bottle was put in the W3 position, and the sippers were inserted to each bottle. The collection trays were placed below the reagent sippers in the dNTP positions. Then the Chlorite Wash was started. After the cleaning wash completed the bottle 250 ml containing the Chlorite solution was removed and the sipper was rinsed. A 250 ml bottle labelled with water wash was filled 250 ml 18 M Ω water and placed in W1 position. A water rinse was started. After the cleaning process was completed all the attached bottles were removed and the reagent sippers were kept.

2.9.3 18 MΩ water wash

Each bottle labelled with water wash cleaning was rinsed with 100 ml 18 MΩ water 3 times. From the Ion PGM touchscreen, 18 MΩ water cleaning option was selected from the Clean menu. Following the instructions that appeared on the touchscreen, a cleaning chip (old chip) was placed in the chip clamp. All wash and reagent bottles attached to the PGM were removed and only the reagent sippers were kept. The 250 ml of was filled with 18 MΩ water and placed in W1 position and attached to a clean sipper. An empty 2 l bottle was placed in W2 position and an empty 250 ml bottle was placed in W3 position and the sippers were attached to each bottle. The collection trays were placed below the reagent sippers in the dNTP positions. The Water wash was started. When the process finished all the bottles and sippers were removed from the W1, W2 and W3 positions. The reagent sippers and collection trays were kept in place. The ION PGM was ready for the initialization.

2.10 Initialization of the Ion PGM™ System

The dNTP stock solutions were removed from the freezer and the tubes were placed in a cold rack. The tank pressure for the nitrogen gas cylinder was checked.

2.10.1 Preparation of the Wash 2 Bottle

The 2 l bottle (wash 2) was washed with 200 ml 18 MΩ water 3 times and 500 µl of 100 mM NaOH was prepared by diluting 50 µl of 1 M NaOH in 450 µl of nuclease-free water. The bottle was filled with 18 MΩ water to the line found in the bottle (around 2 L). The Ion PGM™ Hi-Q™ Sequencing W2 Solution was added to the 2 L (wash 2) bottle. A volume of 70 µl was added to Wash 2 bottle. The bottle was closed and was mixed by inverting the bottle 5 times.

2.10.2 Preparation of the Wash 1 and Wash 3 Bottles

The 250 ml bottles (W1 and W3) were rinsed with 50 ml 18 MΩ water. 350 µl of freshly prepared 100 mM NaOH was pipetted to the W1 wash bottle. Ion PGM™ Hi-Q™ Sequencing W3 Solution was added to the 50 ml line found on the bottle.

2.10.3 Starting the Initialization of Ion PGM™

From the Ion PGM touchscreen, the Initialization option was selected. A new long-gray sipper was attached to the cap in W2 position and the prepared W2 bottle was attached and the cap was tightened. Two gray short sippers were attached to W1 and W3 position. The prepared W1 and W3 bottles were placed in their proper position and attached to the new sippers, and then caps were tightened.

2.10.4 Preparation the 50 ml Reagent Bottles with dNTPs

Four new 50 ml tubes were labelled with the four dNTPS labels as dCTP, dATP, dTTP, dGTP. The four dNTP stock solutions tubes were vortexed and briefly centrifuged. With separate filtered tips 20 µl of each dNTPs were added to the corresponding tube. Between each tube gloves were changed. Blue short sippers were attached into each dNTPs port. According to the dNTPs labels, each 50 ml reagent tube was attached to its correct position and were tightened. The pH was adjusted during the initialization step and a green Passed was appeared when the pH was in the correct target. With the green Passed the ION PGM was ready for the chip sequencing.

2.11 Data analysis

The sequencing PGM data were viewed after the logging into the server and the Data window was selected to view the run results. The Sequencing data were generated from all chips were analysed in Torrent Suit version 5.0.4. The run summary reports for all the chips were viewed and also were exported as PDF files that provided information about the run quality such as the ISP density, ISP summary run and the read length. Also, the plugin HID_SNP Genotyper version 5.2.2 was used. The plugin is a software plug-in tool analyses the samples which are barcoded in the run and looks and locates the genotypes at the position given in the hotspots file. Following the HID SNP Genotyper user guide, the plugin for the Precision Ancestry Panel the Admixture Prediction – AISNPs and Population Likelihoods – AISNPs panels were selected. For the targeted regions the PrecisionID_AncestryPanel_targets.bed file was selected and for the hot regions the PrecisionID_AncestryPanel_hotspots.bed was selected. The Admixture Prediction and Population Likelihoods results were analyzed. Also the plugin was used to analyse the Identity Panel samples and to run the plugin for the Precision Identity 1000 Genomes – IISNPs panels and Y Haplogroup Prediction – Y SNPs Panel

were selected. For the targeted regions the PrecisionID_IdentityPanel_targets.bed file was selected and for the hot regions the PrecisionID Identity-Panel_hotspots.bed was selected. Both Y Haplogroup Prediction and 1000 Genomes results were analyzed. Allele coverage sample table containing the genotypes for the samples sequenced were also studied.

2.12 Statistical analysis

2.12.1 SRTUCRURE Analysis

All the Excel files including the Genotypes for the 300 Qatari samples were exported from the Torrent Server. The results were further analysed using STRUCTURE 2.3.4 software (Pritchard et al., 2010). It was used to estimate the population structure based on the SNP data. This software used a Bayesian probabilistic clustering approach to estimate the population of individuals on the basis of genetic data.

2.12.2 Arlequin Analysis

The Exact test for Hardy –Weinberg equilibrium (p -value), observed heterozygosity (H_o), expected heterozygosity (H_e) and pairwise F_{ST} values and estimation of linkage disequilibrium (LD) were assessed using Arlequin Software version 3.5.1.3 (Excoffier & Lischer, 2010).

2.12.3 PowerStats Analysis

PowerStats V1.2 software Microsoft Excel Workbook template (Tereba, 1999) was used for statistical calculations as allele frequencies and other forensic parameters Discrimination Power (DP), Match of Probability (MP) and Polymorphic Information Content (PIC), Typical Paternity Index (TPI) and Power of Exclusion (PE). Also, for more explanation of the Hardy –Weinberg equilibrium, pairwise F_{ST} , linkage disequilibrium (LD), and forensic parameters see Appendix 3.

CHAPTER: 3 Evaluation of Precision ID Identity Panel in Qatar population by massively parallel sequencing using the Ion Torrent Personal Genome Machine (PGM)

3.1 Introduction

Autosomal short tandem repeats (STR) markers analysed by PCR and capillary electrophoresis (CE) represent the gold-standard for forensic DNA analysis and they are considered as a reliable tool for identification (Al-Awadi et al., 2014; Shrivastava et al., 2016; Tan et al., 2018). Today, paternity and genetic identification in forensic casework are typically carried out by testing a set of 16-24 STR markers. The STRs used are standardised to a large degree to be common between many of the kits; this standardisation has been driven by legislative requirements in the US for the original 13 combined DNA index system (CODIS) core STR markers that grew to an expanded set of 20 STRs (Hares, 2015), and several commercial kits are available that contain these STRs (Oostdik et al., 2014; Ludeman et al., 2018). In the European Union the European Standard set (ESS) of loci was a subset of the CODIS 13 loci (D3S1358, TH01, D21S11, D18S51, vWA, D8S1179 and FGA); the expanded ESS introduced 5 STR loci (D1S1656, D2S441, D10S1248, D12S391 and D22S1045) in addition to the original 7 STR loci (Schneider, 2009)– these were also present in the expanded CODIS loci. Single nucleotide polymorphisms (SNPs) have some advantages over STRs in that they have a lower mutation rate, can be typed with smaller amplicons and can also provide additional genetic information, as parental lineage determination, biogeographical ancestry and phenotypical traits assessment (Avila et al., 2019). It is well known that SNPs markers display a relatively high successful rate with degraded DNA samples in comparison with STRs due to short amplicons length, and therefore SNPs typing protocols have been developed for forensic application (Turchi et al., 2019). However, SNPs use in forensic DNA analysis for identification purposes have been limited due to the restricted capacity of simultaneous typing of different markers in a single run presented by CE method (Avila et al., 2019). The large amounts of DNA (10–200 ng) per sample per experiment needed for the high-throughput platforms which is

not available in forensic casework (Børsting et al., 2009b). Also, the most significant disadvantage of SNP typing is that the limited number of alleles (typically two) for each SNP locus limits or prevents reliable mixture interpretation (Butler et al., 2007).

MPS provides the possibility to detect several hundred markers simultaneously, including different kinds of markers, and also allows multiple samples to be analysed in a joint sequencing run using sample-tagging DNA barcodes (Li, R. et al., 2018).

While all national forensic DNA database are built on STR polymorphisms the situation is changing with the first MPS DNA profile sent to the French National DNA Database in February 2018 (Laurent, 2018).

The Precision ID Identity panel (initially marketed as the HID Ion Ampliseq Identity Panel) is one of the first panels released for forensic studies for human identification. The panel includes 124 SNPs: 90 autosomal SNPs, 43 of them from the Kidd Panel (Pakstis et al., 2010) and 48 SNPs from the SNPforID Consortium Panel (Phillips et al., 2007b) with 1 SNP shared between them. Additionally, 34 Y upper-clade SNPs are incorporated into the kit (Karafet et al., 2008). The panel is designed to amplify forensic samples with an input of 1 ng. The panel has been optimized to identify genotypes from degraded samples more robustly than is possible with commonly used STR kits. After library construction the amplicons can be sequenced on the Ion Torrent Personal Genome machine (PGM) and/or Ion S5 (Thermo Fisher Scientific, 2016b).

Objectives

- a) To evaluate massively parallel sequencing workflow using the Precision ID Identity Panel within the Forensic Laboratory of Qatar.
- b) To genotype 105 unrelated Qatari individuals from the eight municipalities that makeup the State of Qatar.
- c) To evaluate the sensitivity of the Precision ID Identity Panel.
- d) To study the panel performance in terms of Coverage, the 124 SNP profiles quality, heterozygote balance (Hb) for the 90 autosomal SNPs included in the panel and the noise level (NL) for all SNPs.

- e) To evaluate the forensic statistical parameters from the 105 Qatari data and the forensic parameters generated and their importance in casework analysis with STR standard analysis tool in Qatar forensic DNA lab.
- f) To investigate the Y-haplogroup distribution in the Qatari population.

3.2 Materials and Methods

The materials and methods relevant to this Chapter were described in Chapter 2.

3.3 Results

3.3.1 Sensitivity study

The sensitivity of the Precision ID Identity Panel was tested by using by using a buccal swab from male volunteer that was quantified with real-time PCR (5.57 ng/μl) and then serially diluted to provide five template amounts: 1 ng, 0.5 ng, 0.25 ng, 0.05 ng and 0.01 ng. The dilutions were prepared in triplicate for library preparation with 21 PCR cycles.

Different barcodes (IonXpress_001 - IonXpress_015) were used for the libraries and sequenced on Ion 316 Chips using the Ion PGM instrument.

3.3.1.1 Sensitivity study results

All the experiment results were reviewed using the default HID SNP Genotyper parameters including a minimum coverage of 6 reads. The quality checks were generated for each chip and each sample.

From HID SNP Genotyper the results were reviewed as a sample table, which included the genotype list for all the samples across all hotspots. Also, allele coverage was generated for the experiments as tables and charts. The charts tab showed the coverage across the hotspots whereas the table tab shows the allele details such as genotype, coverage and quality check information (QC check). The sample genotypes flagged according to the HID SNP Genotyper QC filter, which included coverage, percent positive coverage, major allele frequency and the genotype (Table 3.1).

Table 3.1. Data in table showing the HID SNP Genotyper Plug-in quality checks.

Quality Checks	The conditions	QC flags
Coverage	Total coverage at locus is less than twice the standard deviation compared to the mean. Mean and standard deviation are computed independently for autosomal and Y SNPs.	COV
Percent Positive Coverage	If the ratio of coverage from positive strand to negative strand is <30% or >70% this indicates strand imbalance.	PPC
Major allele frequency	The ratio of coverage of major allele to total coverage is: <ul style="list-style-type: none"> • <95% for homozygotes • <35% or >65% for heterozygotes 	MAF
Genotype	The genotype is not valid a No Call expressed as NN.	NOC

Full SNP profiles were collected from samples with DNA input of 1 ng, 0.5 ng, and 0.25 ng. One SNP of the 90 autosomal SNPs dropped out in one of the replicates of 0.05 ng. The three replicates of DNA 0.01 ng led to partial SNP profiles as shown in (Table 3.2).

Table 3.2. Data showing the results of the sensitivity study.

Sample	DNA concentration	# of Autosomal SNPs drop out	#of Y SNPs drop out
Replicate 1	1ng	Full	Full
	0.5 ng	Full	Full
	0.25 ng	Full	Full
	0.05 ng	Full	Full
	0.01 ng	2 drop outs (rs4288409, rs2076848)	9 drop outs (rs2534636, rs9786139, rs372687543, P256, rs2032595, rs20320, rs9341278, rs3911,rs2033003)
Replicate 2	1 ng	Full	Full
	0.5 ng	Full	Full
	0.25 ng	Full	Full
	0.05 ng	1 drop out (rs445251)	Full
	0.01 ng	1 drop out (rs826472)	4 drop outs (rs35284970, rs369616152, rs2032599, rs2032673)
Replicate 3	1 ng	Full	Full
	0.5 ng	Full	Full
	0.25 ng	Full	Full
	0.05 ng	Full	Full
	0.01 ng	3 drop outs (rs2016276, rs1528460, rs722098)	11 drop outs (rs9786184, rs369616152, rs4141886, rs17222573, rs372157627, rs3848982, rs3900, rs2032673, rs2032652, rs13447443, rs2033003)

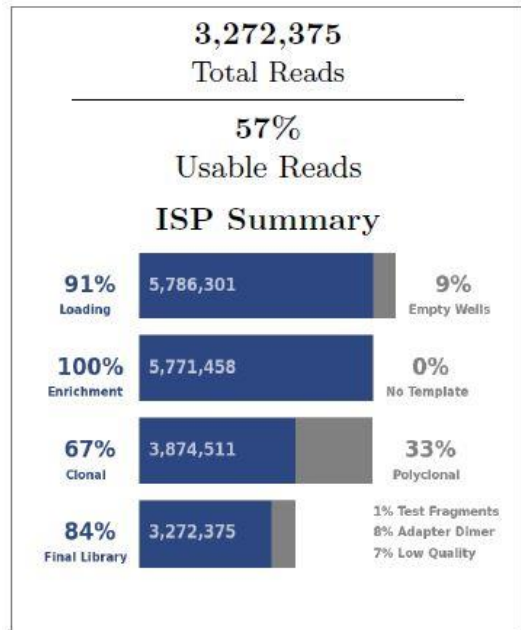
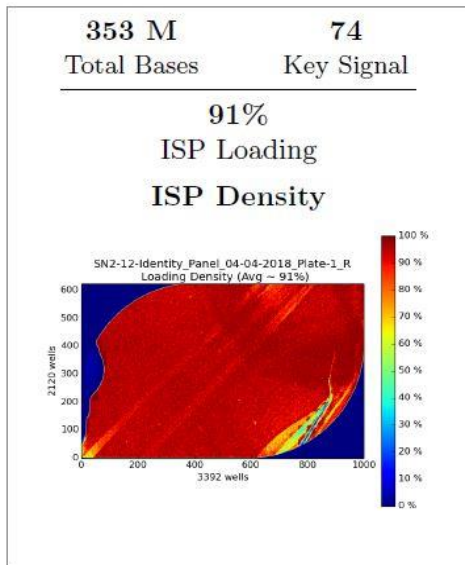
3.3.2 Quality of the sequencing runs

In this study, a total of 105 unrelated Qatari individuals (84 males and 21 females) from the eight municipalities were typed with Precision ID Identity Panel. The library construction was performed using the Precision ID Identity Panel and the Ion Precision ID Library Kit employing 21 amplification cycles. Template preparation and chip loading were performed using the Ion Chef™. Sequencing was performed on the Ion PGM™ sequencer and using four Ion 316 Chip Kit v2 (TFS).

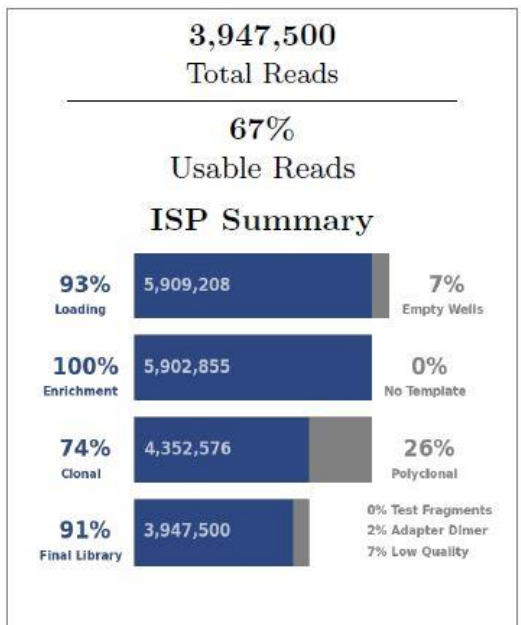
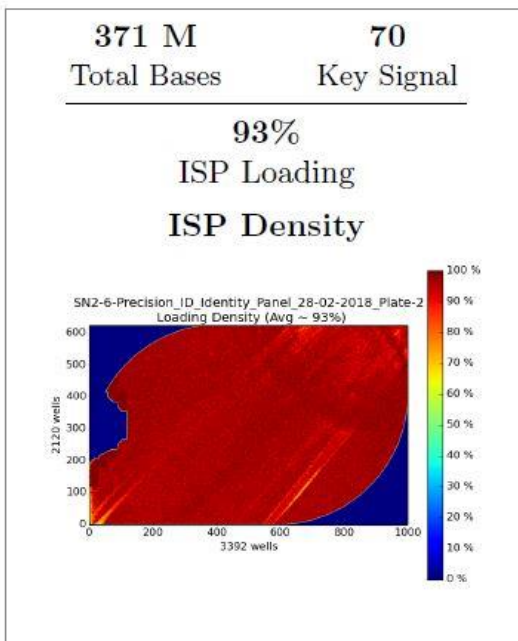
Each run was evaluated using the summary report in the Torrent Suite™ Software which include several parameters. Here, I discussed the parts that explain the quality of the sequencing run, I reviewed the quality metrics for the unaligned reads. The quality metrics for unaligned reads are divided into three categories. First, Quality of chip loading, expressed as the density of Ion Sphere™ Particles (ISPs) loaded onto the chip. It is displayed as a heat map: red indicated a good loading, yellow heat map refers to less ISP, but it is acceptable while the blue and green for the poor loading. The second part of the report showed the ISP summary. The loading percentage always corresponded to the average loading density (ISP density %) which was the number of ISPs in the wells. The percentage of wells that included ISPs while the empty wells indicated the total number of the wells without ISPs, taking into account that the 316 chip contained 6 million wells. The enrichment percentage provided information on how many of ISPs held DNA on them and the quality of ISP (i.e., clonal ISP holding one template) or polyclonal (ISP holding two or more templates), Finally, the percentage of the final library indicated how many reads were in the final library and recorded in the output. The third section of the run report shows a histogram of the read length and it also describes the distribution of the read length of HID Identity panel. See Appendix 4 for the metric and their description.

Here, the chip loading density were good for all the chips: 91% for chip#1, 93% for chip#2 and chip#3 and 94% for chip#4; the ISP summary which explained the quality of the ISPs that were loaded onto the chips (Figure 3.1) and the read length histogram which described the distribution of the read length of HID Identity panel. The mean read lengths for all the 4 chips were: 108 bp, 94 bp, 123 bp and 127 bp.

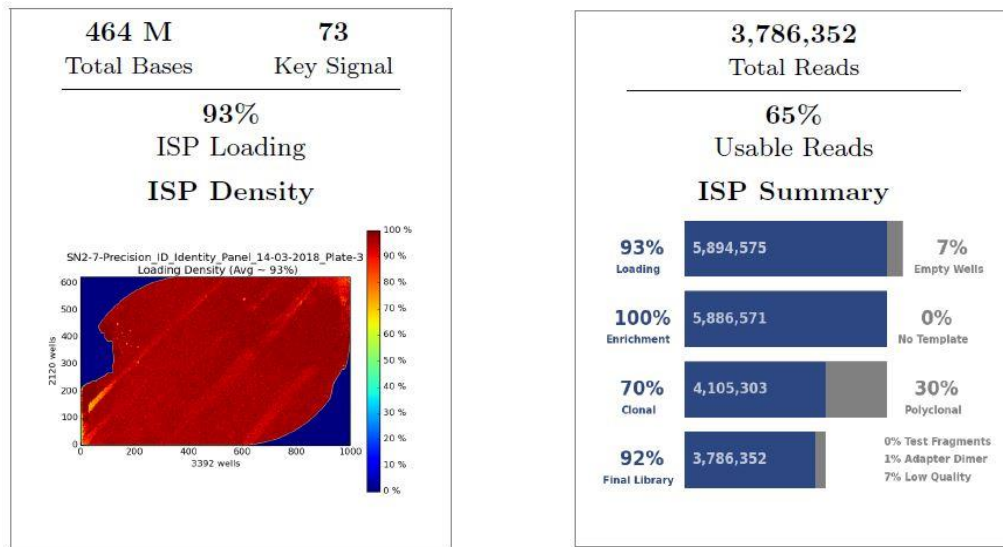
Chip#1



Chip#2



Chip#3



Chip#4

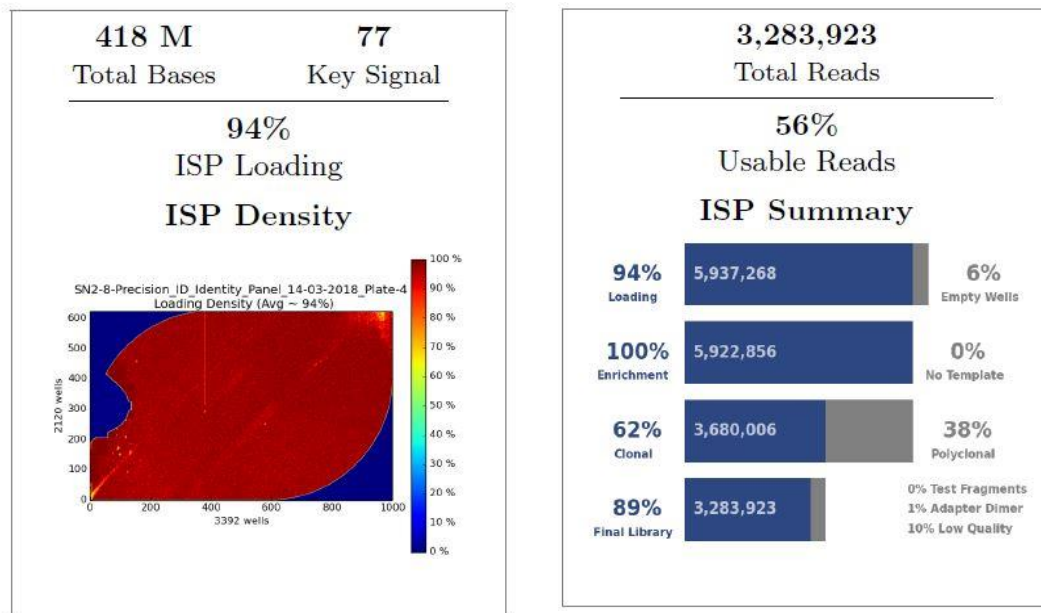


Figure 3.1 The different panels showing the ISP density percentage and ISP summary for Chip#1, Chip#2, Chip#3 and Chip#4 that was taken from summary report for each chip.

3.3.3 Sequence Coverage

The average mean of depth obtained from the coverage analysis report is shown for each of the 4 chips in Table 3.3.

Table 3.3. Data showing the average mean depth of sequencing for the 4 Chips.

Chip#	Average Mean Depth
Chip#1	652.59
Chip#2	682.89
Chip#3	672.41
Chip#4	467.44

The average coverage across the 90 autosomal SNPs and 34 Y-SNPs was 754.81 x and 266.77 x, respectively. The minimum coverage value for the 90 autosomal SNPs was 17 x (rs214955) and for the Y SNPs the minimum values were at M479 (0 x) and rs2032599 (12 x) (Table 3.4 and 3.5).

Table 3.4. Data displaying the autosomal SNPs which showed the lowest coverage values among the 105 samples.

SNP name	Number of samples where the SNP displayed the lowest coverage
rs214955	58
rs2342747	16
rs826472	9
rs2046361	6
rs917118	5
rs876724	4
rs717302	3
rs987640	2
rs12997453	2
Total	105

Table 3.5. Data showing the YSNPs with the lowest coverage values among the 84 male samples.

SNP name	Number of samples where the SNP displayed the lowest coverage
M479	65
rs9341278	11
rs2032599	3
rs2032624	3
rs2032631	1
rs2032636	1
Total	84 samples

3.3.4 Performance of the Precision ID Identity Panel

3.3.4.1 Autosomal SNPs profiles; 90 SNPs

Complete 90-SNP profiles were successfully obtained from all the 105 Qatari individuals.

3.3.4.2 Y SNPs profiles; 34 SNPS

From the 84 male samples, there were 75 samples with full Y-SNPs profiles, while the remaining 9 samples generated partial Y-SNPs profiles with dropouts (Table 3.6).

Table 3.6. Data showing the nine samples showed partial Y-SNP profiles.

Sample name	# of SNPs drop outs	SNP ID
S20-BC20-P1	1	M479
S17-BC17-P2	2	rs2032624, rs2032599
S29-BC29-P2	3	rs2032599, rs2032636, rs9341278
S21-BC21-P3	1	M479
S9-BC9-P4	1	M479
S14-BC14-P4	2	rs2032624, rs13447443
S21-BC21-P4	2	M479, rs2032624
S23-BC23-P4	1	M479
S29-BC29-P4	1	rs2032624

3.3.5 Heterozygote balance and noise level

3.3.5.1 Heterozygote balance

The heterozygote balance (Hb) was calculated for all 90 autosomal SNPs as the number of reads for one nucleotide divided by the number of reads for the other nucleotide in the order A, C, G, and T. The lowest median values were seen in 3 SNPs: rs7520386 (0.404953), rs717302 (0.473856), rs917118 (0.443662) (See Figure 3.2 and Appendix 5). Average Hb of all median values for 90 autosomal SNPs was (0.862) with minimum and maximum values of 0.404953 at rs7520386 and 0.94368 at rs251934.

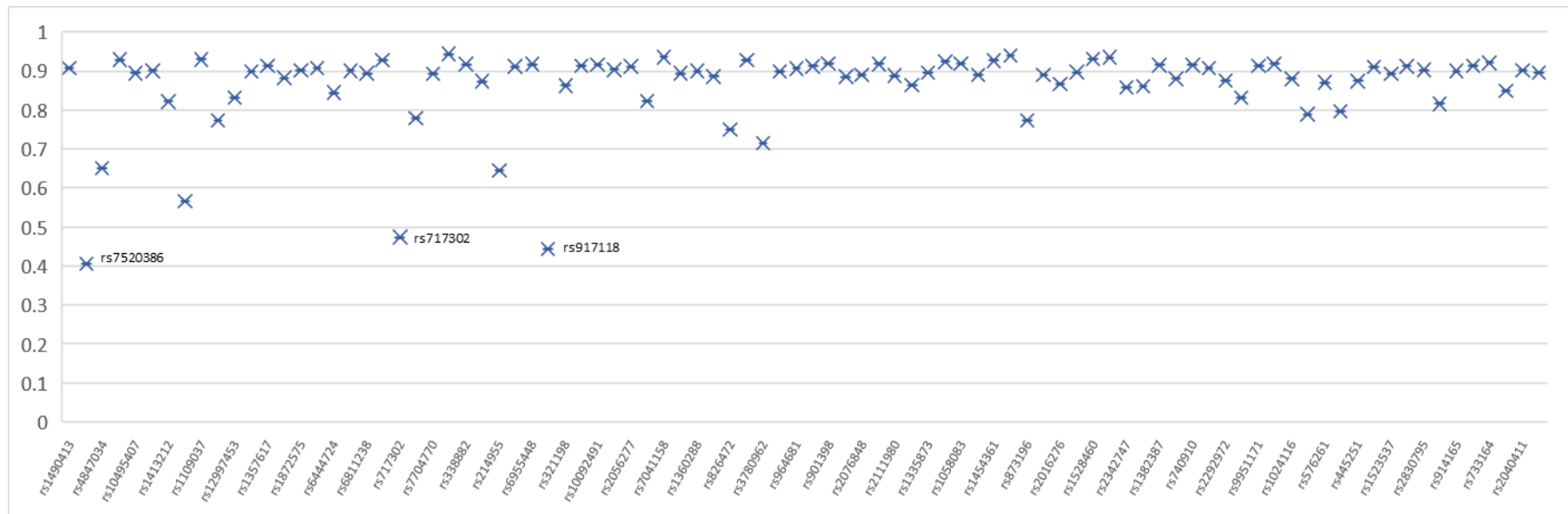


Figure 3.2. Data showing the heterozygote balance values (median) for autosomal SNPs.

3.3.5.2 Noise level

Noise level (NL) results were calculated as the as the coverage of non-alleles divided by the total coverage for each locus (See Figure 3.3 and Appendix 6). In general, median values of the majority of the SNPs approached zero. However, two Y-SNPs performed poorly and the highest median percentage of noise reads was (1.53%) at rs2032636 and (1.69%) at M479.

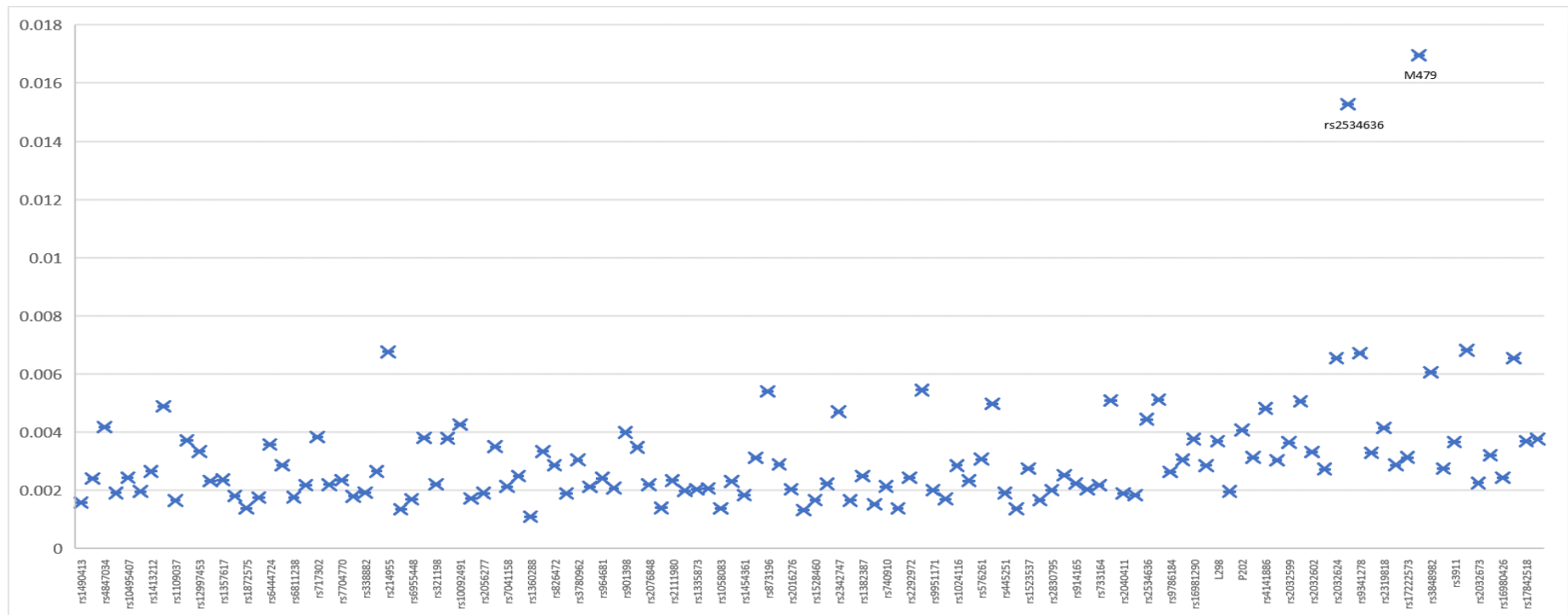


Figure 3.3. Data showing noise level values (median) of the Precision ID Identity panel.

3.3.6 Statistical value of 90 autosomal SNPs of HID Identity SNPs in Qatar Population

All the Excel files including the Genotypes for the 105 Qatari samples were exported from the Torrent Server and used in statistical calculations. For the 90 autosomal SNPs included in the Precision ID Identity Panel, the PowerStats V1.2 software Microsoft Excel Workbook template (Tereba 1999) was used to calculate the allele frequencies, power of discrimination (PD), random match probability (MP), paternity index (PI), probability of exclusion (PE) and polymorphism information content (PIC) calculations. The Exact test for Hardy –Weinberg equilibrium (p -value), observed heterozygosity (H_o), expected heterozygosity (H_e) and pairwise F_{ST} and estimation of linkage disequilibrium (LD) value were assessed using Arlequin (Excoffier et al. 2010).

3.3.6.1 Allele Frequencies of 90 autosomal SNPs of HID Identity SNPs in Qatar Population

The allele frequencies for 90 autosomal SNPs were calculated and are listed in (See Figure 3.4 and Appendix 7).

3.3.6.2 Forensic parameters

The combined matching probability (CMP), and the combined power of exclusion (CPE) for the 90 autosomal SNPs were 7.56748×10^{-37} and 0.999998714. The highest and lowest discrimination power SNP were rs2269355 (PD= 0.664, MP= 0.336) and rs2016276 (PD= 0.364, MP= 0.636), respectively (See Figure 3.5 and Appendix 8).

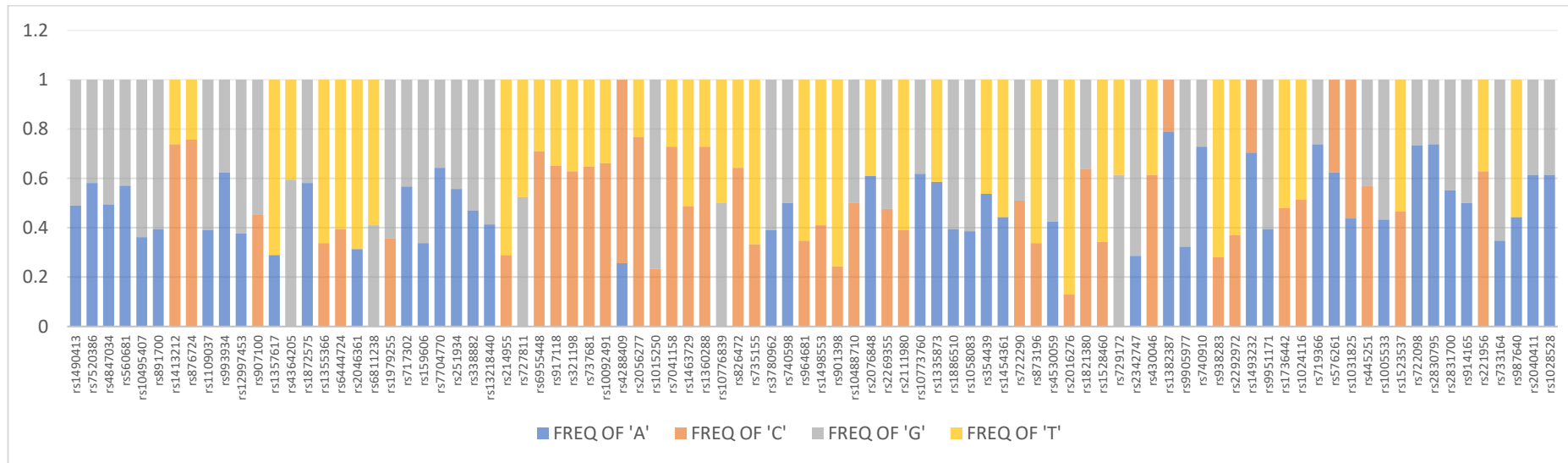


Figure 3.4. Column plot of allele frequencies of 90 autosomal SNPs of Precision ID identity panel in Qatar population

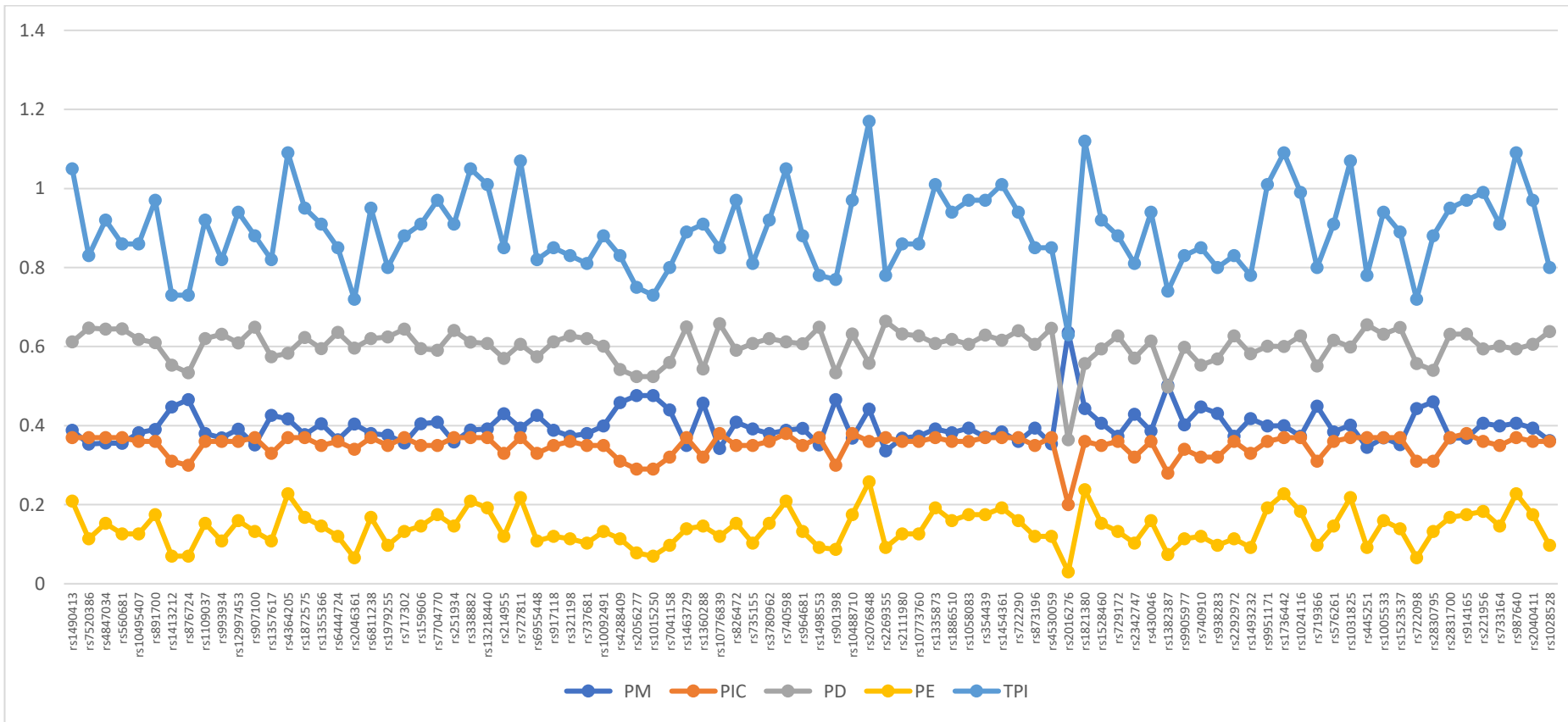


Figure 3.5. Plot graph showing the forensic parameters of 90 autosomal SNPs of The Precision ID Identity panel for 105 Qatari samples, (PM: probability of match, PD: power of discrimination, PE: Power of Exclusion, PIC: polymorphism information content, and TPI: Typical paternity index).

The 105 Qatari individuals studied were divided into three groups (three zones) to calculate the Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD) and F_{ST} genetic distances and Hardy-Weinberg equilibrium (HWE) (see Table 3.7).

Table 3.7. Data showing the three divided populations (groups) from the Qatar population, with a total of 105 samples. The groups were made according to the geographical locations on the map of Qatar.

Population number	Municipality Names -Geographic locations	Sample sizes
P1	Doha - East	20
	Al Wakra - South East	11
P2	Al Rayyan - West	25
	Al Sheehaniya - West	10
P3	Al Shamal – North	5
	Umm Salal - North West	15
	Al Khor and Al Thakira – North East	11
	Al Daayen – North East	8

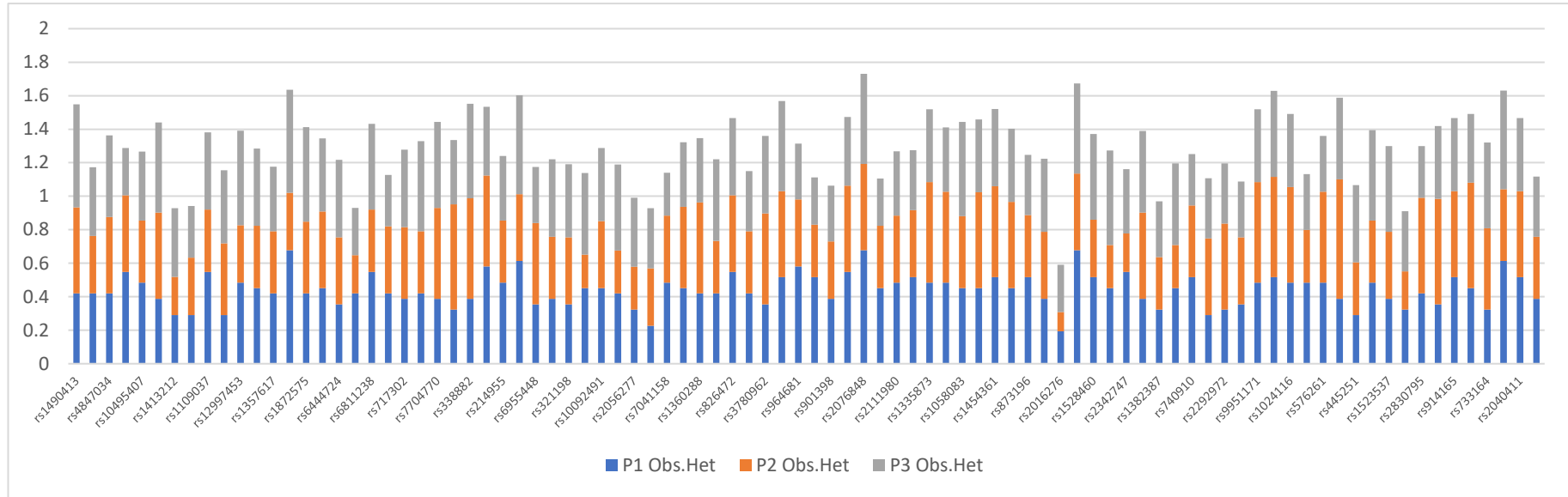
3.3.6.3 Hardy-Weinberg equilibrium

The p - values, observed (Obs.Het) and expected heterozygosity (Exp.Het) values for Hardy-Weinberg equilibrium were performed and summarized. Cells highlighted with yellow represents p -value <0.05 , the minimum observed heterozygosities highlighted with green and the blue colour represents the minimum expected heterozygosities values.

Departures from HWE, at p values <0.05 were observed at 2 SNPs markers in P1 (rs993934) and (rs1821380). Ten SNPs in P2 (rs1413212), (rs2046361), (rs737681), (rs10776839), (rs729172), (rs2342747), (rs9905977), (rs1031825), (rs445251) and (rs722098). In P3 4 SNPs (rs560681), (rs2046361), (rs1498553), (rs2269355). After applying Bonferroni correction 0.0006 (0.05/90) all the loci were in HWE. The

Observed heterozygosity values were in the range 0.19355 (rs2016276) to 0.67742 (rs4364205) in P1, from 0.11429 (rs2016276) to 0.71429 (rs1031825) in P2 and from 0.25641 (rs7041158) to 0.61538 (rs4364205). Whereas the expected heterozygosity values ranged from 0.27499 (rs1382387) to 0.5082 (rs4364205) in P2, from 0.10932 (rs1031825) to 0.50725 (rs445251) and from 0.24542 (rs2016276) to 0.50649 (rs1031825) in P3 (See Figure 3.6 and Appendix 9). The average observed heterozygosities for P1, P2 and P3 are 0.441, 0.425 and 0.431, respectively and the average expected heterozygosities are 0.464, 0.452 and 0.455.

(A)



(B)

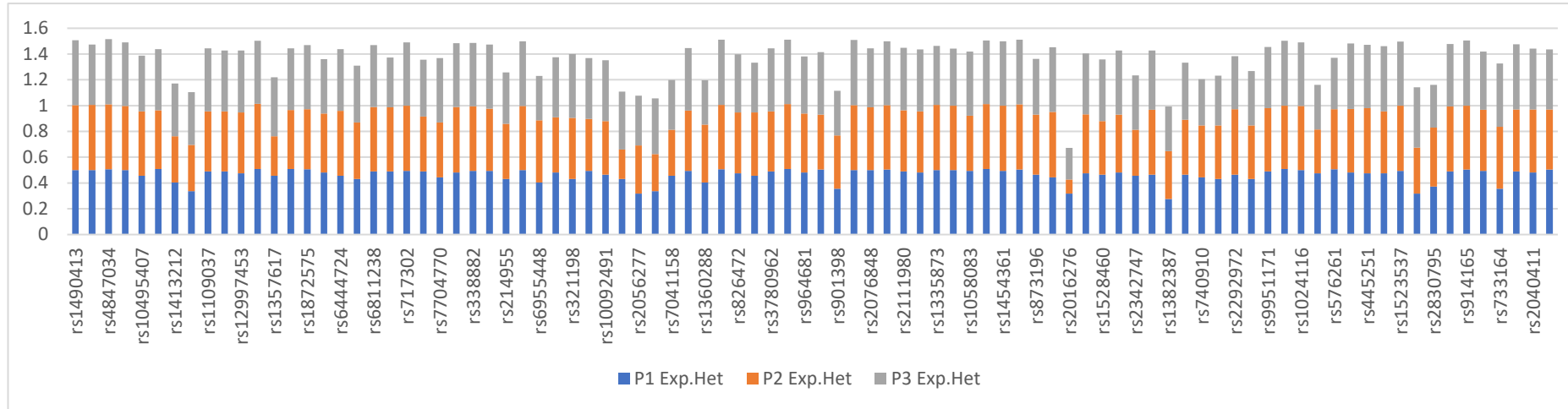


Figure 3.6. (A) and (B) Bar plots of Observed heterozygosities (Obs.Het) and expected heterozygosities (Exp.Het) across 90 autosomal SNPs of the Precision Identity panel typed in 105 Qatari individuals from P1, P2 and P3 groups.

3.3.6.4 Pairwise F_{ST} values

To examine population differentiation among the three Qatari population pairwise F_{ST} values were determined using Arlequin software (see Table 3.8).

All values were negative or extremely low and none were statistically significant; this lack of genetic differentiation is expected given the small size of Qatar and the absence of clear population differences between different regions of the country.

Table 3.8. Data showing the pairwise F_{ST} values between 3 tested population based on 90 Autosomal SNPs included in the panel.

	P1	P2	P3
P1	0.00000		
P2	0.00314	0.00000	
P3	-0.00024	0.00539	0.00000

3.3.6.5 Linkage disequilibrium (LD)

The pairwise linkage disequilibrium (LD) between 90 autosomal loci in each population group of Qatar was also tested (Table 3.9). Significant LD was observed at $p < 0.05$: 384 pairs of loci in the P1, 380 pairs in the P2 and 394 pairs in the P3. However, after Bonferroni correction was applied at $p < 0.00000631$ (0.05 divided by 7921 (89 x 89 loci number of loci involved)) only 1 pair of loci in P3 was found to be remain significant. The association is between locus rs735155 (at Chromosome 10) and locus rs2076848 (at Chromosome 11) (Appendices 10: A, B, and C).

Table 3.9. Data showing the significant linkage disequilibrium at $p < 0.05$ in three populations of Qatar.

SNP#	SNP Names	P1	P2	P3
0	rs1490413	6	2	3
1	rs7520386	3	6	3
2	rs4847034	4	4	6
3	rs560681	5	4	2
4	rs10495407	2	1	1
5	rs891700	3	4	3
6	rs1413212	2	4	3
7	rs876724	3	3	4
8	rs1109037	8	5	6
9	rs993934	2	5	3
10	rs12997453	5	12	5
11	rs907100	9	7	5
12	rs1357617	4	2	2
13	rs4364205	5	5	5
14	rs1872575	3	2	9
15	rs1355366	4	8	6
16	rs6444724	3	3	4
17	rs2046361	7	5	3
18	rs6811238	4	1	1
19	rs1979255	2	2	4
20	rs717302	1	6	3
21	rs159606	3	4	4
22	rs7704770	0	1	2
23	rs251934	4	6	8
24	rs338882	5	6	6
25	rs13218440	3	0	6
26	rs214955	6	6	3
27	rs727811	3	6	4
28	rs6955448	2	2	8
29	rs917118	5	4	4
30	rs321198	5	0	4
31	rs737681	1	6	5
32	rs10092491	2	5	4
33	rs4288409	6	7	3
34	rs2056277	7	6	3
35	rs1015250	7	3	6
36	rs7041158	4	1	4
37	rs1463729	8	3	8
38	rs1360288	1	5	5
39	rs10776839	6	8	1

SNP#	SNP Names	P1	P2	P3
40	rs826472	5	4	3
41	rs735155	5	6	4
42	rs3780962	3	4	3
43	rs740598	1	3	2
44	rs964681	3	5	6
45	rs1498553	5	3	4
46	rs901398	6	3	3
47	rs10488710	5	5	5
48	rs2076848	3	7	6
49	rs2269355	3	8	5
50	rs2111980	3	4	3
51	rs10773760	5	4	6
52	rs1335873	4	3	3
53	rs1886510	5	3	3
54	rs1058083	1	4	1
55	rs354439	7	4	2
56	rs1454361	5	2	6
57	rs722290	5	4	4
58	rs873196	4	2	10
59	rs4530059	5	6	6
60	rs2016276	6	6	5
61	rs1821380	3	8	10
62	rs1528460	4	5	4
63	rs729172	1	1	4
64	rs2342747	5	2	1
65	rs430046	9	6	4
66	rs1382387	6	3	5
67	rs9905977	5	4	4
68	rs740910	3	4	6
69	rs938283	2	5	9
70	rs2292972	8	2	4
71	rs1493232	8	4	4
72	rs9951171	3	4	5
73	rs1736442	3	5	3
74	rs1024116	5	5	6
75	rs719366	4	3	6
76	rs576261	4	2	4
77	rs1031825	4	3	2
78	rs445251	4	7	6
79	rs1005533	3	4	4
80	rs1523537	5	4	5
81	rs722098	5	6	7
82	rs2830795	2	6	7

SNP#	SNP Names	P1	P2	P3
83	rs2831700	3	4	2
84	rs914165	4	3	3
85	rs221956	5	4	2
86	rs733164	5	4	5
87	rs987640	6	6	3
88	rs2040411	7	2	5
89	rs1028528	6	4	5
Total linked loci		384 (192 pairs)	380 (190 pairs)	394 (197 pairs)

3.3.7 Y-haplogroups

The HID SNP Genotyper plug- identifies Y-haplogroups based on the 2014 International Society of Genetic Genealogy (ISOGG) Haplogroup Tree (<http://isogg.org/tree/index14.html>). The plugin displays list of information with identified haplogroup (Table 3.10) and the haplogroup is identified with at least 30% of the Y SNPs should have a valid genotype. (Figure 3.7).

Table 3.10. Data showing the plugin information of the Y Haplogroup predicted [taken from HID SNP Genotyper Plugin (Thermo Fisher Scientific, 2017b)].

Items	Descriptions
Haplogroup	The identified haplogroup.
Profile	The genotypes of the SNPs, provided in the Family Tree DNA (FTDNA) format. Only loci that are in the derived state are represented. Use the profile information to search against third-party tools, such as http://ytree.morleydna.com .
Incompatible loci	Loci that are present in the derived state but are not compatible with the identified lineage.
Y Tree	The Y tree with clades that have been typed for this lineage. Loci that define each clade are color-coded: <ul style="list-style-type: none"> • Red—loci in the ancestral state (Ancestral SNP) • Green—loci in the derived state (Mutant SNP) • Grey—loci that do not have a valid genotype (No data)
Description	A short description of the haplogroup, including details about its origin and geographical distribution.

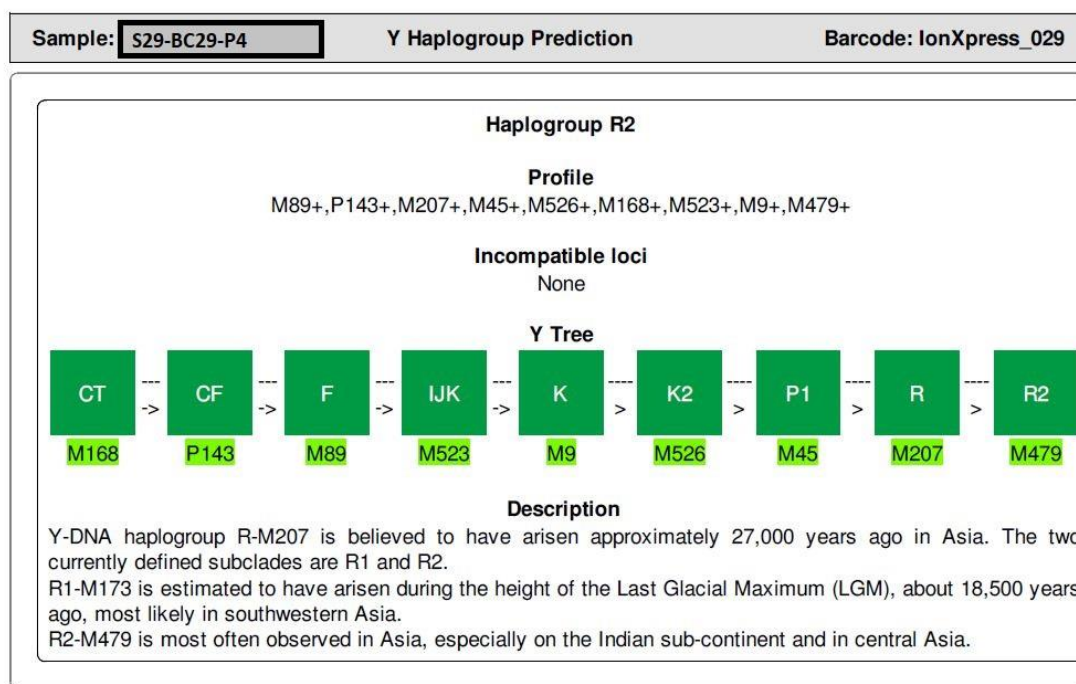


Figure 3.7: A typical example of the information provided for the predicted Y-haplogroup.

Analysis using Precision ID Identity panel was performed in a total 84 Qatari samples, 10 different Y-haplogroups were observed; 60 out of 84 individuals were assigned to J haplogroup, which were distributed in all eight municipalities (Table 3.11). Also, Table 3.12 representing the haplogroup predicted and the mutations that defined the haplogroups from the Y tree.

Table 3.11. Data showing the total of 10 different Y haplogroups were observed among 84 Qatari individuals.

Y- Haplogroups	Doha	Al Rayyan	Umm Salal	Al Shamal	Al Wakra	Al Khor and Al Thakira	Al Sheehaniya	Al Daayen	Total	%
R2	0	2	2	0	0	0	0	0	4	4.76
R1a1	1	0	2	0	4	2	0	0	9	10.71
J	7	11	8	5	6	9	10	4	60	71.43
T	0	1	1	0	0	0	0	0	2	2.38
Q	0	1	0	0	0	0	0	0	1	1.19
F	1	0	0	0	0	0	0	0	1	1.19
E	2	0	0	0	0	0	0	1	3	3.57
B	1	0	0	0	0	0	0	0	1	1.19
G	1	0	0	0	0	0	0	0	1	1.19
L	0	0	0	0	1	0	0	1	2	2.38
Total	13	15	13	5	11	11	10	6	84	100.00

Table 3.12. Data showing the mutation that defined the given Haplogroup.

Haplogroups	Mutations
R2	M479
R1a1	SRY10831.2
J	S35
T	M184
Q	M242
F	M89
E	P171
B	M181
G	M201
L	M20

3.4 Discussion

In the present study, the sequencing performance, forensic parameters, and Y-haplogroup distributions of the Precision ID Identity panel were first investigated in the Qatari population using the Ion Torrent Personal Genome Machine (PGM). For this study, 105 unrelated Qatari individuals from the eight municipalities were typed to investigate the forensic parameters and Y-haplogroup distribution using the Precision ID Identity Panel and the Ion PGM.

The results of the sensitivity study were full SNP profiles seen down to 0.25 ng DNA input using 21 PCR cycles. DNA input of 0.05 ng was also generated full profiles in two replicates and dropout was seen in one replicate in rs445251 (Autosomal SNP). Partial profiles with 11, 5 and 14 dropouts were observed in 0.01 ng replicates. Amplification of 0.01 ng generated partial profiles and showing an approximate 87.8 %, 94.4 % and 84.4 % respectively success rate with total of 90 autosomal SNPs included in the panel, which recovery of (79, 85 and 76) loci. Meiklejohn and Robertson (2017) evaluated the panel and they demonstrated that it is possible to obtain reliable and reproducible genotypes using the Precision ID Identity Panel, when using low quantities (≥ 0.2 ng) of either pure native DNA or forensic type DNA samples. Similarly, Guo et al (2016) reported that full profiles could be obtained from 100 pg input DNA, but the optimal input DNA would be 1 ng–200 pg with 21 PCR cycles. The experiment demonstrated that the capability of generating full and useful profiles below the recommended DNA input.

For the Precision ID Identity panel, the manufacturer's recommended that target average heterozygosity should be $50 \pm 15\%$ (Thermo Fisher Scientific, 2017b). Additionally, approximately 35–45% of all SNPs in the panel for a given sample should be heterozygotes (Buchard et al., 2016). The heterozygote balance Hb for most of the SNPs in the Qatari population, were well balanced with values near to 1; the lowest median values were seen in 3 SNPs: rs7520386 (0.404953), rs717302 (0.473856), rs917118 (0.443662). The lowest Hb values for the SNPs rs7520386 and the rs917118 were (0.154884) and (0.2619050). In addition, these SNPs performed poorly and similar results were reported in other studies (Avila et al., 2019; Buchard et al., 2016; Li, R. et al., 2018; Liu, J. et al., 2018). One of the QC check flags provided by the

Coverage Analysis plugin is MAF (Major Allele Frequency). The three loci flagged due to MAF, which is defined as the ratio of coverage of major allele to total coverage is <95% for homozygotes and <35% or >65% for heterozygotes genotypes which indicate to an allele imbalance.

For the Identity Panel, the manufacturer's recommended, 32 samples could be examined on Ion 316, requiring mean depth 738X for autosomal SNPs and 236 X coverage for Y-clade SNPs to obtain a minimum coverage of 300x and 150x for 97% of A-SNPs and Y-SNPs under the condition of 80% chip loading and 60% usable reads (Thermo Fisher Scientific, 2017b). In this study, although, the number of samples per chip was 30, which is fewer than the 32 samples recommended, some loci received low coverage values. The chip loading densities were: 91% for chip#1, 93% for chip#2 and chip#3 and 94% for chip#4 and the usable reads which is the percentage of library ISPs that pass the polyclonal, low quality, and primer-dimer filters for the four chips were: 57%, 67%, 65% and 56% respectively. This correlates with the good performance with full profiles recovered from all the 105 Qatari individuals with the minimum coverage value for the 90 autosomal SNPs of 17 x in rs214955. This SNP was also reported by Zhang et al. (2015) as one of the loci with low coverage while the partial profiles were seen in the Y-SNPs which were also reported in this study the most problematic sites: M479, rs2032599 as explained below. Buchard and co-worker (2016) suggested that the low number of samples per run ensures sufficiently high coverage and reduces the number of samples with incomplete profiles that need to be re-typed. Suggestions might be made to design future experiments to run a lower number of samples per run, especially in case of samples that contained low concentrations of DNA.

Meiklejohn and Robertson (2017) and Buchard et al. (2016) noted that the coverage for Y-clade SNPs was less than that obtained for autosomal SNPs which is not unexpected given Y-clade SNPs have only one copy compared to the two copies for each autosomal SNP; in addition. female samples have no coverage.

Generally, most of the Noise Level (NL) median values near zero across all 124 SNPs. Two Y-SNPs performed poorly and the highest median percentage of NL reads were (1.53%) at rs2032636 and (1.69%) at M479. The minimum and maximum percentage of

NL reads at rs2032636 and M479 were (0.22%, 9.83%) and (1.27%, 7.14%) respectively, and these findings were similar to Avila et al. the two Y-SNPs performed poorly than the autosomal markers with larger median value (Avila et al., 2019).

In this study the NL values of a few samples reached above 10% for some loci (rs7520386, rs826472, rs1024116, rs2032631). A similar result was reported by Buchard et al. (2016) and Liu, et al. (2018) with NL values of a few samples reach above 10%. It was clear that the performance of the autosomal SNPs was better than the Y-SNPs and that through the full genotypes which were collected and also the acceptable median values of the noise analysis. In contrast, partial SNP profiles and signals of noise were observed in the Y-SNPs.

This preliminary study showed that the panel performed well and as there is no manufacturer's recommendation for calculating Hb and noise level, this is why the study took this approach. Since the findings from this preliminary study will be used in future casework analysis, SNPs that performed poorly whether autosomal and/or Y-SNPs listed above should be carefully analyzed when analyzing the low source of DNA samples or in mixture analysis.

Qatar population was divided into three groups of population to study HWE, F_{ST} and LD. The average observed heterozygosities for P1, P2 and P3 are 0.441, 0.425 and 0.431, respectively and the average expected heterozygosities are 0.464, 0.452 and 0.455. Departures from HWE, at p value <0.05 was seen in 16 loci. However, no significant ($p > 0.05$) deviation from HWE was detected in the distribution of the 90 autosomal SNPs after Bonferroni correction.

Only one SNP pair remained and displayed high significant linkage equilibrium ($p = 0.0000$) even after Bonferroni correction and that is probably due to the relatively small sample size, as the markers are on different chromosomes.

The combined matching probability (CMP) and the combined power of exclusion (CPE) for the 90 autosomal SNPs were 7.5674×10^{-37} and 0.999998714. The highest and lowest discrimination powers were in rs2269355 (PD= 0.664, MP= 0.336) and rs2016276 (PD= 0.364, MP= 0.636), respectively.

3.5 Conclusion

In summary, the results have demonstrated that the Precision ID Identity panel is a reliable and robust tool and the forensic statistical parameters obtained proved to be powerful for human identification purposes in the Qatari population. The high number of SNPs (90 autosomal SNPs) included in the panel allowed to generate promising results and to reach to high discrimination power. The CMP for a 15 STR markers including all 13 CODIS core markers of the Identifiler Plus in Qatari population 1.90×10^{-20} , lower than CMP provided by Precision ID Identity Panel. These findings demonstrate that autosomal SNP set included in the Precision ID Identity shows sufficient levels of polymorphism and informative to be of value for Qatari population. Before implementing in casework samples from other nationalities in Qatar, it may be necessary to ensure that match probabilities/likelihood ratios do not overstate the strength of the evidence. The combination of autosomal and Y-SNPs could provide valuable information which can help to use the panel along with routine STRs casework analysis and paternity testing in some kind of complex cases with the presence of mutations or in cases with complex pedigrees.

CHAPTER: 4 Massively parallel sequencing of forensic samples using the Precision ID Identity panel on the Ion Torrent PGM.

4.1 Introduction

Chapter 3 is related to the evaluation of the Precision ID Identity panel for use on reference samples. The data obtained in Chapter 3 demonstrated that the methodology was working effectively in the Forensic DNA Laboratory in Qatar and that the sensitivity and general performance were similar to other reported studies.

However, the generation of profiles from reference samples is not particularly challenging and can be done routinely with standard STR analysis. Before considering whether the SNP testing would be a valuable tool in forensic casework it needs to be tested on forensic samples to gauge its performance in comparison to the current technology.

One of the main challenges and limitations when analyzing forensic samples of low quantity or that are highly degraded is that either partial or no profile is produced (Jordan & Mills, 2021; Turchi et al., 2019). SNPs markers in the Precision ID Identity panel have been designed with small amplicons. In this Chapter, the panel performance and suitability for real forensic samples was tested using real forensic samples collected from several cases that had been previously analyzed with different STR kits.

The aim of this part of the study is to evaluate the performance and suitability of the Precision ID Identity Panel on real casework samples selected from a range of casework.

Objectives

- a) To evaluate the performance and reliability of the Precision ID Identity panel with real casework samples that previously extracted with different extraction protocols used in the Qatari forensic laboratory.
- b) To compare the performance of the SNP analysis with conventional STR analysis

4.2 Materials and Methods

The study was carried out on 60 casework samples from 46 cases, of which 55 samples were also analysed by Precision ID Ancestry panel (Chapter 6). The samples were collected between 2005 and 2018 and comprised samples that had both partial and full genotypes. The samples were previously genotyped with one or more of Identifiler®, Identifiler® Plus, MiniFiler™ and GlobalFiler®; some samples had also been amplified with Yfiler® and Yfiler Plus® (Applied Biosystems™).

4.3 Results

Sixty casework samples were selected from different biological materials including hair, cells, semen, saliva, blood, touch, teeth and bone. The samples were available as extracted DNA. Library construction was performed using the Precision ID Identity Panel and the Ion Precision ID Library Kit (TFS). The samples with DNA concentration of 1 ng and above were amplified with 21 cycles whereas for samples with less than 1 ng available 26 cycles of amplification were used. An Ion Chef™ was used for template preparation and chip loading. Templated amplicons were sequenced using Ion Torrent PGM instrument (TFS) on two 316 v2 Chips (Table 4.1).

Table 4.1. Table showing the total number of casework samples collected for the Precision ID Identity panel experiment.

Sample Types	Chip5*	Chip6
Hair	-	1
Cells	2	8
Semen	3	2
Bone	1	1
Blood	11	5
Saliva	11	5
Touch	1	7
Tissue	1	1

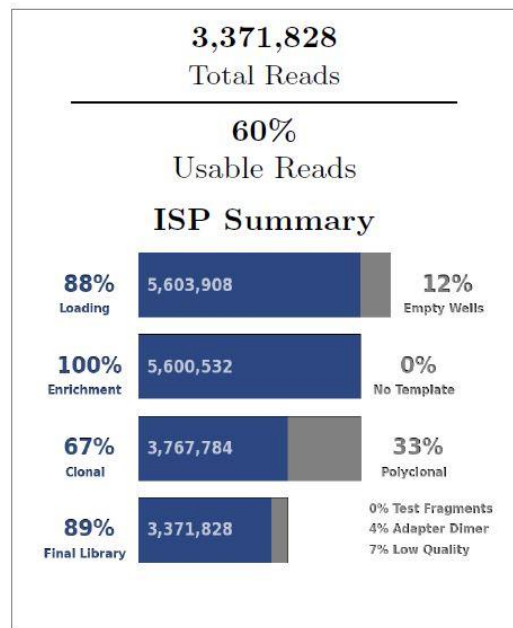
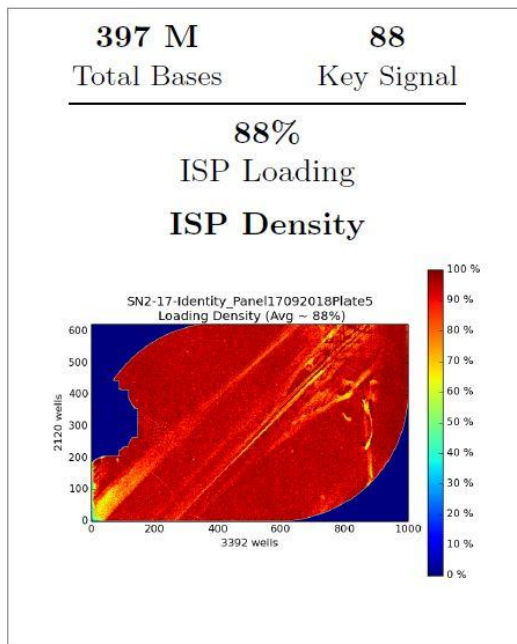
* Chip numbering followed on from the numbering in Chapter 3.

4.3.1 Quality of the sequencing runs

The quality for the two runs was evaluated through the output reports generated by Torrent Suite™ Software 5.4.0, through Summary and Quality control metrics. The first tier of report was primarily focussed on ISP Density, ISP Summary, and Read Length that were used to evaluate every run.

The chip loading density was 88% for Chip#5 and 81% for chip#6. The ISP summary which detailed the quality of the ISPs that were loaded onto the two chips is shown in Figure 4.1. The read length histogram describes the distribution of the read length of HID Identity panel. The mean and median read lengths for all the two chips were (118 bp, 109 bp), (119 bp, 97 bp), respectively.

Chip #5



Chip #6

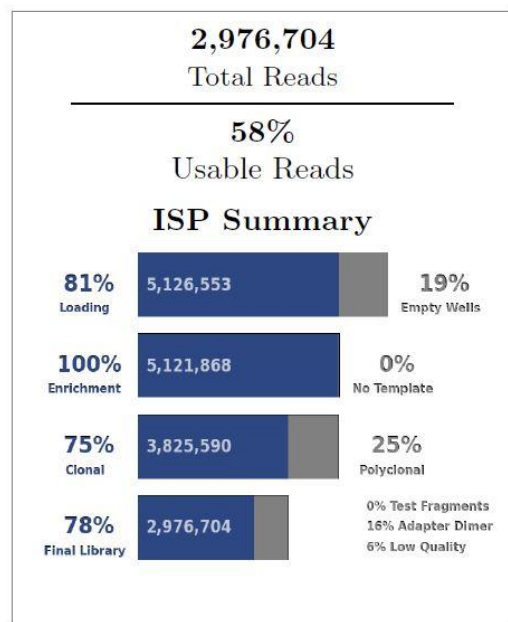
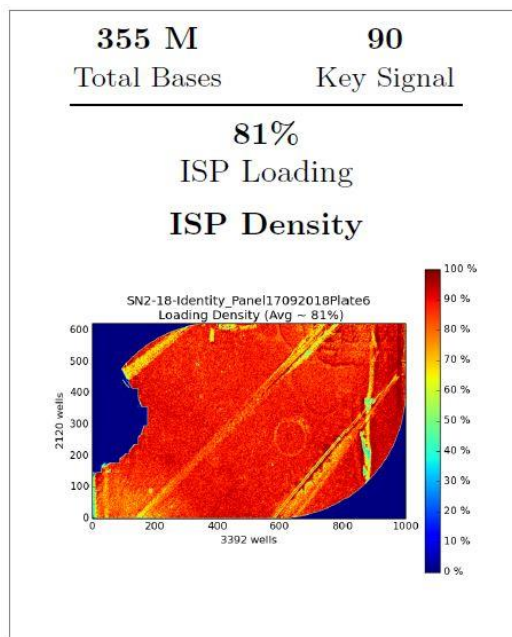


Figure 4.1. Chip diagrams showing the ISP density percentage and ISP summary for Chip#5 and Chip#6 taken from summary report for each chip.

4.3.2 Sequence Coverage

The average mean of depth was obtained from the coverage analysis report as shown in Table 4.2 for two 316™ Chips.

Table 4.2. Data showing the average mean depth for Chip#5 and Chip#6.

Chip#	Average Mean Depth
Chip#5	613.23
Chip#6	129.53

In the current experiment, the analyses were carried out using the default setting: as minimum coverage 6 reads. The average coverage across the 90 autosomal SNPs and 34 Y-SNPs were 448.54x and 200.68 x respectively. Tables 4.3 and 4.4 show the SNPs with lowest coverage values that were observed within the samples. Figure 4.2 shows the maximum and median values for each SNP; the average median value was 369.5x with range from 60x (rs722290) to 640x (rs1058083) among autosomal SNPs, whereas the Y-SNPs the average median value was 147x with range from 40x (rs2032599) to 288.5x (rs9786184).

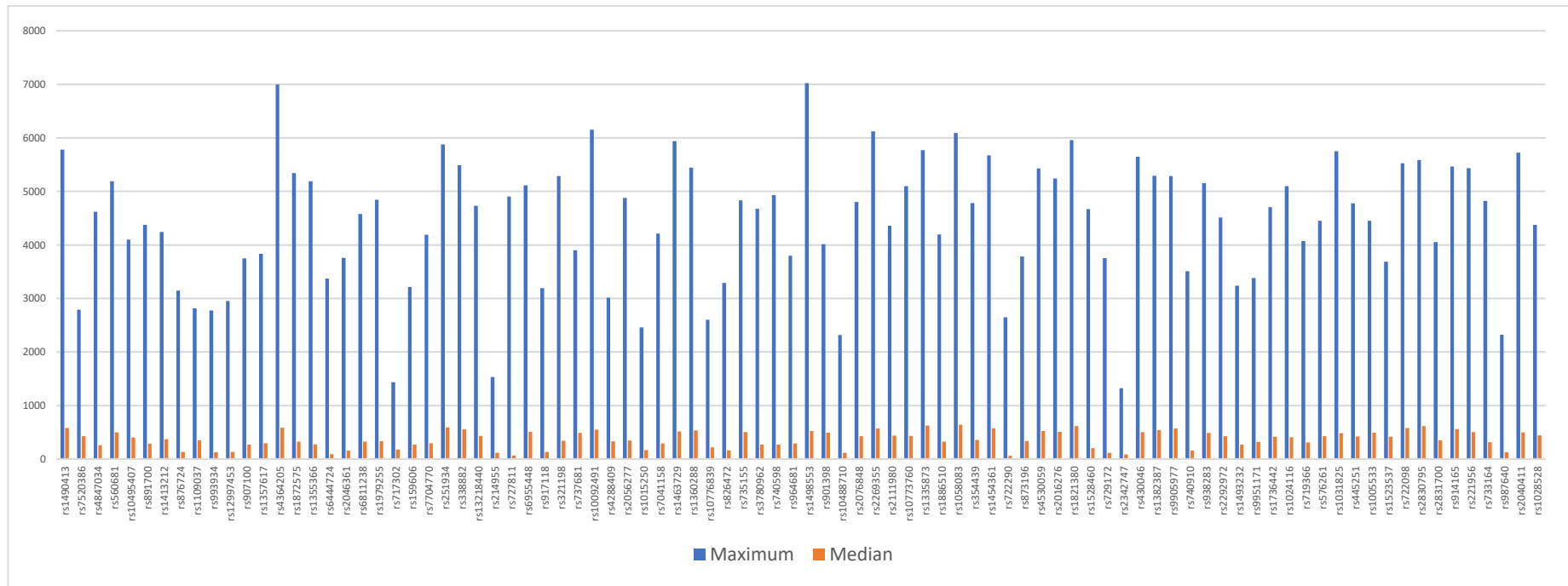
Table 4.3. Table showing the autosomal SNPs which showed the lowest coverage values.

SNP name	Number of samples	Chip#
rs214955	10	5
rs722290	9	6
rs987640	7	5,6
rs917118	5	5
rs1490413	5	6
rs2342747	4	5
rs826472	4	5,6
rs727811	3	5,6
rs993934	2	5,6
rs876724	1	5
rs354439	1	5
rs10776839	1	5
rs10488710	1	5
rs717302	1	5
rs729172	1	6

Table 4.4. Table showing the Y-SNPs with the lowest coverage values.

SNP name	Number of samples	Chip#
M479	25	5,6
rs2032599	12	5,6
rs2032624	4	5,6
rs16981290	4	5,6
rs2534636	3	6
rs2032652	2	5,6
rs9341278	2	6
rs2032636	1	6
rs2032658	1	6

(A)



(B)

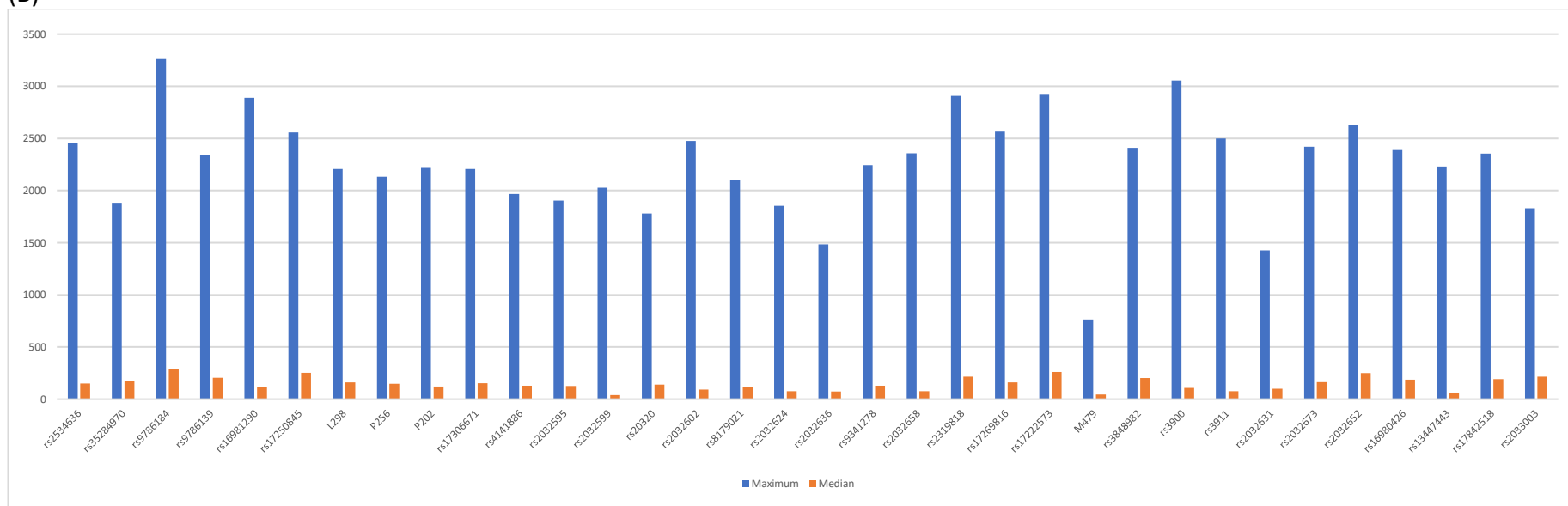


Figure 4.2. (A) and (B) Column chart showing maximum and median of the coverage values for each SNP.(A) Showing the 90-Autosomal SNPs. (B) Showing the 34- YSNPs included in the Precision ID Identity panel resulted from casework samples.

4.3.3 Performance of the Precision ID Identity Panel

4.3.3.1 Full SNP profile

Full profiles of 124 SNPs were successfully generated from 27 casework samples representing several types of biological evidence (Table 4.5). The full profiles were seen in samples with a concentration of 1 ng and higher and also in 6 weak samples with less than 1 ng input DNA (highlighted in orange). The yellow row in the Table 4.5 represents a full profile generated from the lowest DNA concentration (i.e. 0.01 ng).

Table 4.5. Data showing the full SNP profiles were seen in 27 casework samples

Sample#-Barcode#-Plate#	Year of Analysis	Quantification values (ng/μl)*	Sample Types and information, if available	STR results
S1-BC1-P5	2005	10.34	Cells collected from teeth cleaning twig "Miswak"- Unknown	PPlus, Re-amp-GF/FP
S5-BC5-P5	2015	17.02	Bone- Unknown	IDP, Re-amp-GF/FP
S6-BC6-P5	2018	25.79	Blood stain blade knife-Libyan	GF/FP
S7-BC7-P5	2017	12.25	Touch (swab from bullets)- Yemeni	GF/FP
S8-BC8-P5	2017	0.98	Cells from bandage	IDP/FP
S10-BC10-P5	2016	27.90	Saliva	IDP/FP
S11-BC11-P5	2018	46.12	Blood stain	GF/FP
S12-BC12-P5	2016	11.52	Semen stain (Sperm fraction)	IDP, Re-amp-GF/FP
S13-BC13-P5	2016	11.75	Blood stain	IDP, Re-amp-GF/FP
S15-BC15-P5	2018	2.22	Blood stain	GF/FP
S16-BC16-P5	2018	50.15	Blood stain	GF/FP

Sample#-Barcode#-Plate#	Year of Analysis	Quantification values (ng/μl)*	Sample Types and information, if available	STR results
S17-BC17-P5	2017	13.72	Saliva	IDP/FP
S18-BC18-P5	2016	49.34	Saliva	IDP/FP
S19-BC19-P5	2016	76.88	Saliva	IDP/FP
S20-BC20-P5	2016	40.17	Saliva	IDP/FP
S21-BC21-P5	2016	30.62	Saliva	IDP/FP
S23-BC23-P5	2017	116.48	Blood stain	GF/FP
S24-BC24-P5	2017	26.52	Semen stain (Sperm fraction)	GF/FP
S25-BC25-P5	2017	23.55	Saliva	GF/FP
S27-BC27-P5	2018	2.9	Saliva from cigarette	GF/FP
S30-BC30-P5	2018	9.9	Semen stain	GF/FP
S1-BC1-P6	2018	0.32	Cells from tissue paper	GF/FP
S2-BC2-P6	2016	0.12 [^]	Saliva. Chewing gum	IDP, Re-amp-GF/FP
S9-BC9-P6	2015	0.01 [^]	Tissue	IDP, Re-amp-GF/FP
S25-BC25-P6	2018	0.5	Blood stain	GF/FP
S27-BC27-P6	2017	0.5	Blood stain	GF/FP
S28-BC28-P6	2017	0.5	Blood stain	GF/FP

*All DNA sample concentration was adjusted to 1 ng and volumes based on quantity of DNA. [^] The maximum volume of extracts (6 μl) was added in case of some samples yielded concentrations of DNA below that value. FP- Full profile, PP- Partial profile, GF- Global Filer, IDP- Identifiler Plus and PPlus- Profiler plus. Re-amp- sample re-amplified (depending on the extract volume available some samples were reamplified with GlobalFiler®).

4.3.3.2 Partial SNP profiles

Partial SNP profiles from the autosomal SNPs with genotypes called as (NN) and/or (-) and from the YSNPs with (N) and/or (-). As shown below in Table 4.6, 28 samples generated partial SNP profiles. Four samples generated profiles with 123 SNPs which represents 99.19% profile recovery (cells highlighted in green). Also, partial profiles with (2-5) SNPs dropping out were generated from 6 samples (cells highlighted in blue). Partial profiles with 6-11 SNPs dropping out were generated from three samples and are highlighted in orange colour. Partial profiles with 24 and 23 SNPs missing genotype data were seen in two samples (highlighted in pink). The partial profiles collected from the remaining samples had different rates of drop-out, with the highest drop-out seen were in two samples with 121 and 118 SNPs not profiled, the samples had DNA quantification values of 0.19 ng and 0.03 ng respectively (highlighted in yellow).

Table 4.6. Data showing partial SNP profiles were seen in 28 casework samples.

Sample#- Barcode#- Plate#	Quantificat ion values (ng/μl)*	Sample Types and Year of analysis	#SNP missed the Genotype		STR results	Percentage of SNP profile recovery
			Autosom al SNPs	Y SNPs		
S9-BC9-P5	17.21	Saliva-2017	FP	1	IDP/FP	99.19%
S22-BC22-P5	8.22	Saliva-2016	FP	1	IDP/FP	99.19%
S26-BC26-P5	118.10	Blood-2017	FP	1	GF/FP	99.19%
S24-BC24-P6	0.5	Saliva from cigarettes-2018		1	GF/FP	99.19%
S2-BC2-P5	16.77	Tissue-2005	FP	3	PPlus /FP	97.58%
S3-BC3-P5	1.13	Blood-2016	FP	5	IDP/FP	95.96%
S4-BC4-P5	4.56	Blood-2016	FP	5	IDP/PP (Y drop out)	95.96%
S29-BC29-P5	8.99	Blood-2017	FP	5	GF/PP (Y drop out)	95.96%
S17-BC17-P6	0.05	Semen stain. Sperm fraction- 2017	4	1	IDP/FP	95.96%
S21-BC21-P6	0.5	Saliva from Cigarette-2017		4	GF/FP	96.77%
S14-BC14-P5	6.3	Saliva-2018	2	7	GF/FP	92.74%
S14-BC14-P6	0.07^	Cells. Swab from a cup-2018	1	10	GF/PP	91.12%
S20-BC20-P6	0.25	Semen stain - 2018	5	2	GF/FP	94.35%
S29-BC29-P6	0.5	Blood stain-2017	7	17	GF/FP	80.64%
S30-BC30-P6	0.5	Cells from towel- (cut for extraction) 2018	6	17	GF/FP	81.45%
S28-BC28-P5	0.70	Blood-2017	30	30	GF/PP (Y drop out)	51.61%

Sample#- Barcode#- Plate#	Quantificat ion values (ng/μl)*	Sample Types and Year of analysis	#SNP missed the Genotype		STR results	Percentage of SNP profile recovery
			Autosom al SNPs	Y SNPs		
S3-BC3-P6	0.53	Touch. Swab from steering wheel- 2017	51	30	GF/FP	34.67%
S4-BC4-P6	0.06^	Touch. Swab from Steering wheel- 2018	21	18	GF/FP	68.54%
S7-BC7-P6	0.52	Cells. scratch collected from inside hat-2018	37	29	GF/FP	46.77%
S8-BC8-P6	0.61	Cells collected from traditional male headdress. -2018	68	30	GF/FP	20.96%
S10-BC10-P6	0.20	Cells. Swab from rope-2017	80	31	IDP/FP	10.48%
S12-BC12-P6	1.03	Cells collected from Shirt-2005	27	18	PPlus/FP	63.70%
S13-BC13-P6	0.01^	hair root-2018	50	32	GF/PP	33.87%
S15-BC15-P6	0.2	Touch. Swab from handle screw- 2018	51	26	GF/PP	37.90%
S11-BC11-P6	0.01^	Touch. Swab from gun trigger-2017	83	34	GF/PP	5.64%
S6-BC6-P6	0.19	Swab from knife blade-handle- 2018	84	34	GF/FP	4.83%
S22-BC22-P6	0.19	Saliva from Cigarette-2017	87	34	GF/FP	2.41%
S23-BC23-P6	0.03^	Touch-swab from watch-2018	85	33	GF/FP	4.83%

* All DNA sample concentration was adjusted to 1 ng and volumes based on quantity of DNA. ^ The maximum volume of extracts (6 μl) was added in case of some samples yielded concentrations of DNA below that value. FP- Full profile, PP- Partial profile, GF- Full profile, PP- Partial profile, GF- Global Filer, IDP- Identifiler Plus and PPlus- Profiler Plus.

4.3.3.3 No SNP profiles

Out of 60 casework samples, 5 samples failed to generate any profile (Table 4.7).

Table 4.7. The results illustrating the 5 samples that failed to generate any profile.

Sample#- Barcode#-Plate#	Years of Analysis	Quantification values (ng/ μ l)*	Sample Types	STR results
S5-BC5-P6	2018	0.19	Touch-swab from knife blade-handle	GF/FP
S16-BC16-P6	2016	0.7	Saliva-chewing gum	IDP/FP
S18-BC18-P6	2016	0.01 [^]	Bone	GF/PP
S19-BC19-P6	2016	0.16	Cells swab from a glove	IDP/FP
S26-BC26-P6	2017	0.94	Blood stain	GF/FP

* All DNA sample concentration was adjusted to 1 ng and volumes based on quantity of DNA. [^] The maximum volume of extracts (6 μ l) was added in case of some samples yielded concentrations of DNA below that value. FP- Full profile, PP- Partial profile, GF- Full profile, PP- Partial profile, GF- Global filer and IDP- Identifiler Plus.

4.3.4 Y-haplogroups

The HID SNP Genotyper plugin identifies Y-haplogroups based on 2014 International Society of Genetic Genealogy (ISOGG) Haplogroup Tree (<http://isogg.org/tree/index14.html>). The output exported from the plugin includes information with identified haplo-group which are the sample profile in the form of Family Tree DNA (FTDNA) format and only loci that are in the derived state are represented. The Plug-in displayed the loci with three different colours: red (Ancestral SNP) loci in the ancestral state, green (Mutant SNP) loci in the derived state and grey (No data) loci that do not have a valid genotype. Also, the plugin provides a description including haplo-group details as a summary of its origin and geographical distribution. Analysis using Precision ID Identity panel that includes 34-YSNPs was performed in a total 60 casework samples. Table 4.8 shows the 8 different Y-haplo-groups which were observed in this study.

Table 4.8. Data showing the Y- haplogroup obtained form 43 casework samples.

Y Haplogroups	Number of samples
Haplogroup E	9
Haplogroup H1	1
Haplogroup R1a1	5
Haplogroup J	17
Haplogroup IJ	4
Haplogroup Q	1
Haplogroup L	5
Haplogroup T	1
*No haplogroups found	17
Total	60

Out of 60 selected samples, 17 samples the plugin reported them as No haplo-groups found (Table 4.9).

Table 4.9. Table showing 17 casework samples reported with "No haplo-groups found".

Sample names	DNA Concentration, year of analysis & sample type	Total SNPs dropped		Total SNPs genotyped	STR results, case report information
		Autosomal	Y		
S5-BC5-P5	17.02 ng- 2015-Bone	-	34	90	FP, Unknown Female Sample
S28-BC28-P5	0.70 ng -2017-Blood stain	30	30	64	PP (Y drop out), No Amelogenin Y-allele detected-Male, Nepalese
S3-BC3-P6	0.53 ng-2017-swab from steering wheel	51	30	43	FP, Unknown male Sample
S5-BC5-P6	0.19 ng-2018- Touch-swab from knife blade-handle	90	34	-	FP, Unknown male Sample

Sample names	DNA Concentration, year of analysis & sample type	Total SNPs dropped		Total SNPs genotyped	STR results, case report information
		Autosomal	Y		
S6-BC6-P6	0.19 ng-2018- Touch-swab from knife blade-handle	84	34	6	FP, Unknown male Sample
S7-BC7-P6	0.52 ng -20158- Cellular material collected from inside hat	37	29	58	FP, Unknown male Sample
S8-BC8-P6	0.61 ng-2018- Cellular materials collected from headcover for men	68	30	26	FP, Unknown male Sample
S10-BC10-P6	0.20 ng-2017-Swab from rope	80	31	13	FP, Unknown male Sample- Nepalese
S13-BC13-P6	0.01 ng-2018 Hair root	50	32	42	PP, Unknown male Sample
S15-BC15-P6	0.2 ng-2018-Touch, swab from screw	51	26	47	FP, Unknown male Sample
S16-BC16-P6	0.16 ng-2016- Saliva-chewing gum	90	34	-	FP, Unknown male Sample
S18-BC18-P6	0.01 ng-2016-Bone	90	34	-	PP, Unknown male Sample
S19-BC19-P6	0.16 ng-2016-Cells swab from a glove	90	34	-	FP, Unknown male Sample
S22-BC22-P6	0.19 ng-2017-Saliva from Cigarette	87	34	3	FP, Unknown male Sample
S23-BC23-P6	0.03 ng-2018-Touch-swab from watch	85	33	3	FP, Unknown male Sample
S26-BC26-P6	0.94 ng-2017-Blood stain	90	30	-	FP, Unknown male Sample

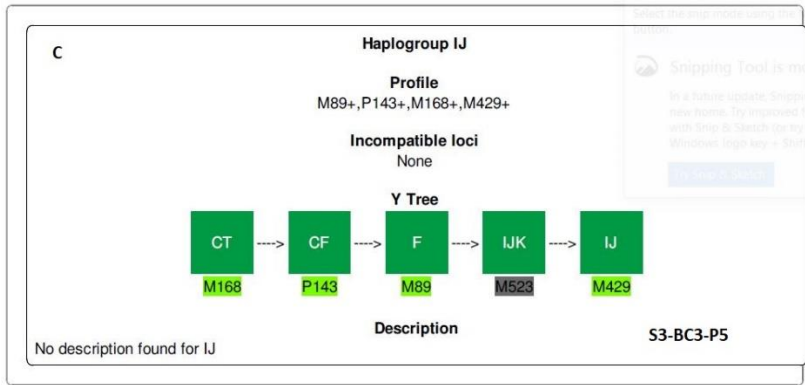
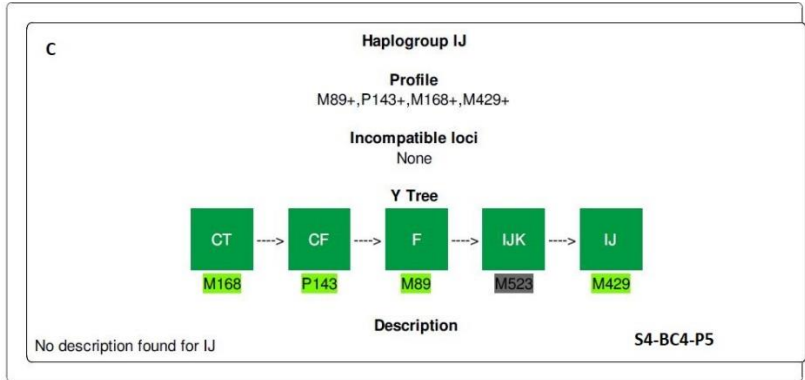
As mentioned above, the 60 samples were previously analysed and their results recorded in the case report file. A group of samples were selected mentioned below which, when analysed some issues were observed during the interpretation of the STR results.

Samples S3-BC3-P5 and S4-BC4-P5:

The two samples were previously amplified with Identifiler® Plus. Sample S4-BC4-P5 was analysed and reported with a drop out at locus AMEL in Y allele and then the profile uploaded to the national DNA database. Later, S3-BC3-P5 was received and the analysis ended with a full profile and an imbalance in the X and Y alleles. When sample S3-BC3-P5 entered to the national DNA database a match to S4-BC4-P5 in 15 loci and dropout of the amelogenin Y allele was found. The Precision ID Identity panel results showed partial SNP profiles collected from both samples with 5 YSNPs missing, however the Y haplogroup was concordant and predicted as Haplogroup IJ (see Figure 4.3).



Sample name	rs2040411	rs1028528	rs2534636	rs3528497	rs9786184	rs9786139	rs1698129	rs1725084	rs3726875	P256	rs3696161	rs1730667
S3-BC3-P5	AG	GG	C	C	C	-	-	-	-	N	T	A
S4-BC4-P5	AG	GG	C	C	C	-	-	-	-	-	T	A



S4-BC4-P5	RMP from 1000 genomes - IISNPs	Barcode: IonXpress_004																								
	<table border="1"> <thead> <tr> <th>Population Name</th> <th>Geo Region</th> <th>RMP</th> <th>D</th> </tr> </thead> <tbody> <tr> <td>South Asia</td> <td>South Asia</td> <td>1.43E-37</td> <td></td> </tr> <tr> <td>East Asia</td> <td>East Asia</td> <td>7.26E-38</td> <td></td> </tr> <tr> <td>Europe</td> <td>Europe</td> <td>1.15E-39</td> <td></td> </tr> <tr> <td>America</td> <td>America</td> <td>2.57E-40</td> <td></td> </tr> <tr> <td>Africa</td> <td>Africa</td> <td>2.39E-42</td> <td></td> </tr> </tbody> </table>	Population Name	Geo Region	RMP	D	South Asia	South Asia	1.43E-37		East Asia	East Asia	7.26E-38		Europe	Europe	1.15E-39		America	America	2.57E-40		Africa	Africa	2.39E-42		
Population Name	Geo Region	RMP	D																							
South Asia	South Asia	1.43E-37																								
East Asia	East Asia	7.26E-38																								
Europe	Europe	1.15E-39																								
America	America	2.57E-40																								
Africa	Africa	2.39E-42																								
S3-BC3-P5	RMP from 1000 genomes - IISNPs	Barcode: IonXpress_003																								
	<table border="1"> <thead> <tr> <th>Population Name</th> <th>Geo Region</th> <th>RMP</th> <th>D</th> </tr> </thead> <tbody> <tr> <td>South Asia</td> <td>South Asia</td> <td>1.43E-37</td> <td></td> </tr> <tr> <td>East Asia</td> <td>East Asia</td> <td>7.26E-38</td> <td></td> </tr> <tr> <td>Europe</td> <td>Europe</td> <td>1.15E-39</td> <td></td> </tr> <tr> <td>America</td> <td>America</td> <td>2.57E-40</td> <td></td> </tr> <tr> <td>Africa</td> <td>Africa</td> <td>2.39E-42</td> <td></td> </tr> </tbody> </table>	Population Name	Geo Region	RMP	D	South Asia	South Asia	1.43E-37		East Asia	East Asia	7.26E-38		Europe	Europe	1.15E-39		America	America	2.57E-40		Africa	Africa	2.39E-42		
Population Name	Geo Region	RMP	D																							
South Asia	South Asia	1.43E-37																								
East Asia	East Asia	7.26E-38																								
Europe	Europe	1.15E-39																								
America	America	2.57E-40																								
Africa	Africa	2.39E-42																								

Figure 4.3. Diagrams showing (A) the STR profiles of AmpFISTR Identifiler Plus system of S3-BC3-P5 and S4-BC4-P5, (B) partial YSNPs profile collected from the sample, (C) Y haplogroup predictions and (D) RMP results of these samples.

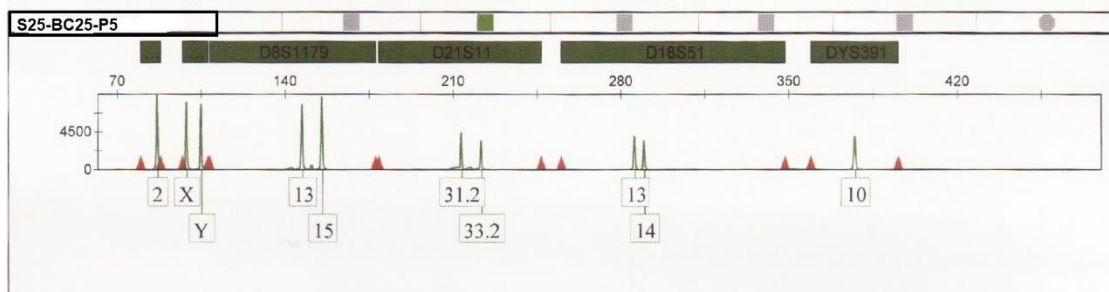
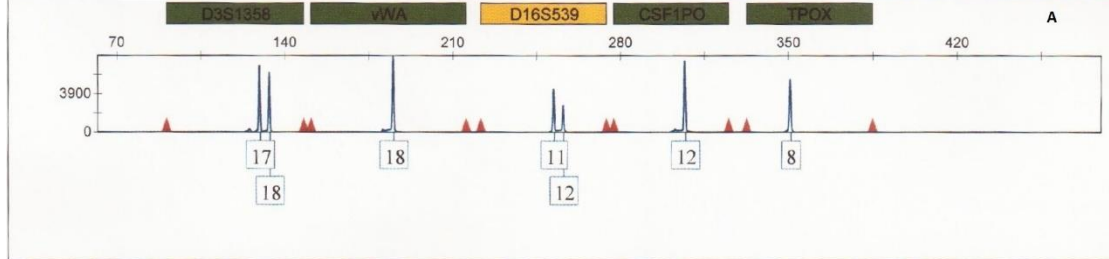
Sample S24-BC24-P5 and S25-BC25-P5:

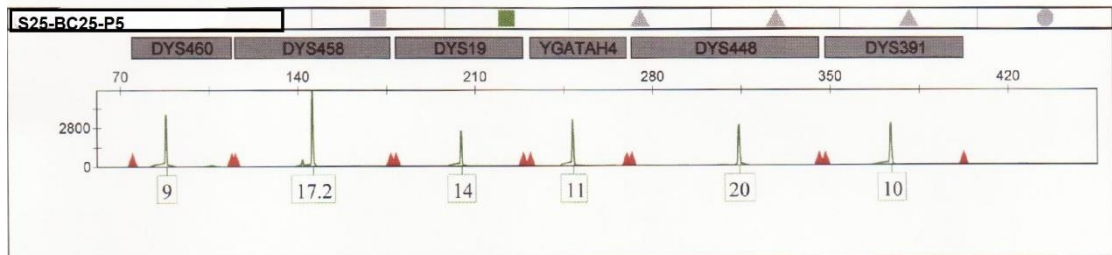
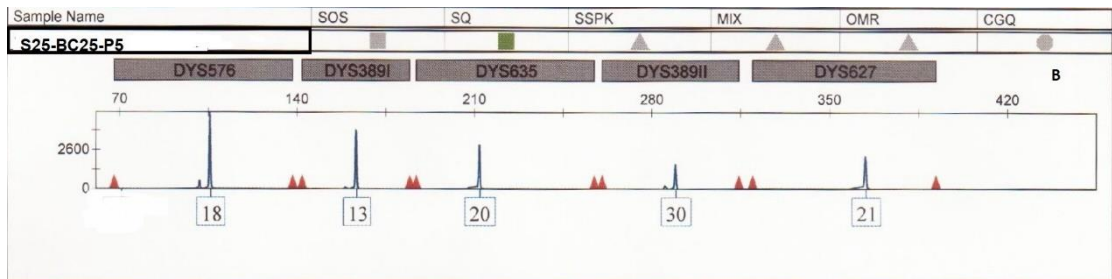
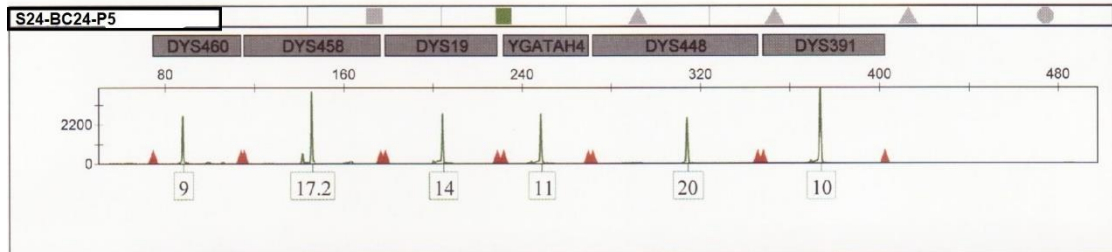
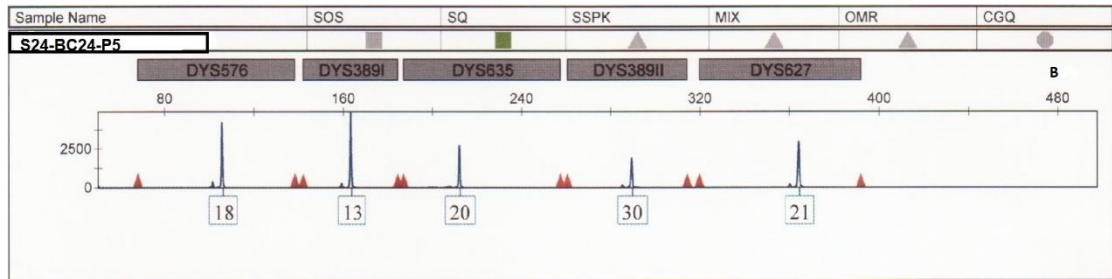
A full match was found between the samples in only the YSTR profiles. In addition, results collected from this experiment showed full 124 SNP profiles and complete YSNPs profiles; the Y haplogroup was Haplo-group J (see Figure 4.4).

Sample Name	SOS	SQ	SSPK	MIX	OMR	CGQ
-------------	-----	----	------	-----	-----	-----

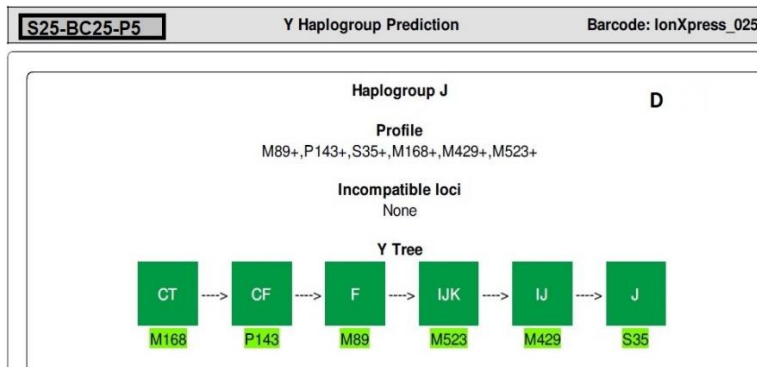
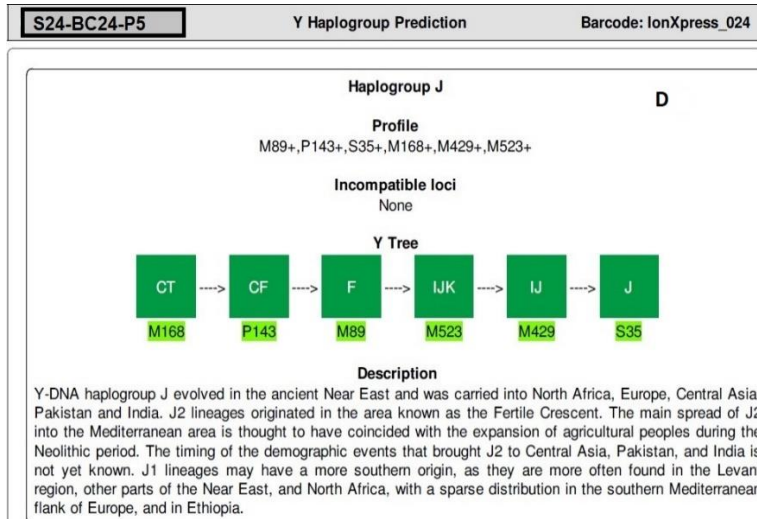


Sample Name	SOS	SQ	SSPK	MIX	OMR	CGQ
-------------	-----	----	------	-----	-----	-----





Sample name	rs1299745	rs907100	rs1357617	rs4364205	rs1872575	rs1355366	rs6444724	rs2046361	rs6811238	rs1979255	rs717302	rs159606	rs7704770	rs251934
S24-BC24-P5	AA	CG	TT	GG	AG	CC	CT	AT	GT	CG	AG	AG	AA	AG
S25-BC25-P5	AA	GG	AT	GT	GG	CT	CT	AT	GG	CG	AG	AA	AG	AA
Sample name	rs2534636	rs3528497	rs9786184	rs9786139	rs1698129	rs1725084	rs3726875	P256	rs3696161	rs1730667	rs4141886	rs2032595	rs2032599	rs20320
S24-BC24-P5	C	C	C	G	C	C	T	G	T	A	A	T	T	G
S25-BC25-P5	C	C	C	G	C	C	T	G	T	A	A	T	T	G



S25-BC25-P5 RMP from 1000 genomes - IISNPs Barcode: IonXpress_025

Population Name	Geo Region	RMP	E
Europe	Europe	2.51E-39	
America	America	1.00E-40	
South Asia	South Asia	8.20E-44	
Africa	Africa	1.56E-44	
East Asia	East Asia	8.66E-45	

S24-BC24-P5 RMP from 1000 genomes - IISNPs Barcode: IonXpress_024

Population Name	Geo Region	RMP	E
Europe	Europe	5.33E-38	
America	America	1.92E-39	
South Asia	South Asia	3.94E-40	
East Asia	East Asia	5.60E-42	
Africa	Africa	8.46E-43	

Figure 4.4. Diagrams showing (A) part of the electropherogram (epg) generated from S24 and S25 using GlobalFiler, (B) part of the electropherogram (epg) generated from S24 and S25 using YFiler Plus, (C) part of the genotype generated from these samples using 124 SNPs included in the Precision ID Identity Panel and (D) Y haplogroup prediction and (E) RMP results of these samples.

Sample S28-BC28-P5:

The sample was a blood stain and the GlobalFiler profile generated from 23 loci with a drop out of the Y allele. The results obtained from the Identity panel experiment were a partial SNP profile with drop out of 30 autosomal SNPs and 30 YSNPs and thus the Y haplo-group not detected (see Figure 4.5).

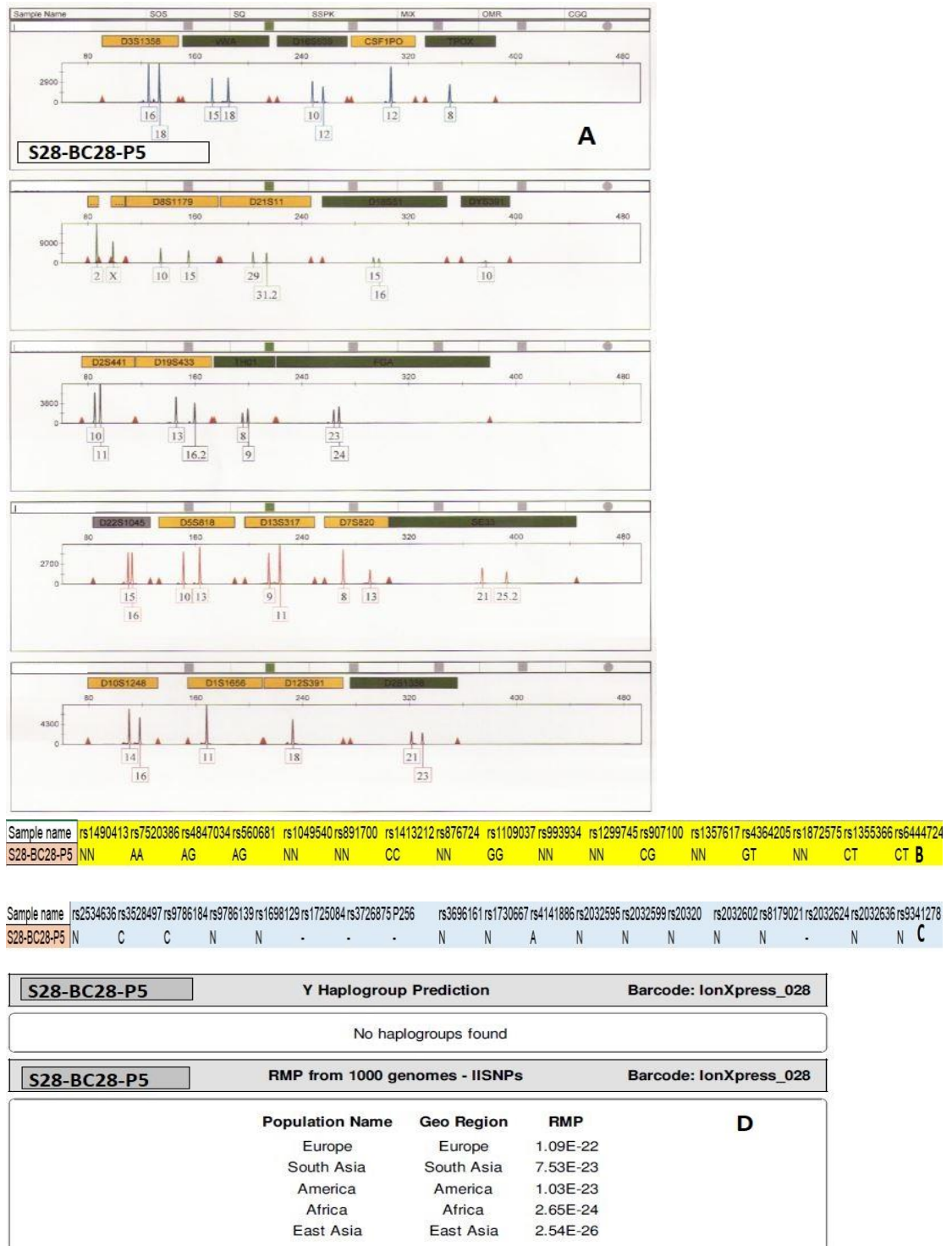


Figure 4.5. Diagrams showing (A) the electropherograms generated from sample S28 using GlobalFiler, (B) part of the partial autosomal SNPs profile, (C) the partial part of the Y-SNPs and (D) Y haplogroup prediction and RMP results of sample S28-BC28-P5.

4.4 Discussion

In the present study, the Precision ID Identity panel from Thermo Fisher Scientific was evaluated. This experiment was applied to 60 casework samples were available as extracted DNA, of which 55 samples were previously analysed by Precision ID Ancestry panel (see Chapter 6). The samples were collected from real cases between 2005 and 2018 that varied from routine stains to challenging samples. The samples had been previously amplified with Identifiler®, Identifiler® Plus, MiniFiler™, GlobalFiler®, some samples were amplified with Yfiler® and Yfiler Plus® (Applied Biosystems™) and had yielded both partial and complete profiles.

Currently, available studies that have investigated the existing Precision ID Identity Panel with 124 SNPs are limited. Guo et al. (2016) reported that the HID ID Identity panel had the ability to work with routine and difficult casework samples and the full SNP genotype was typically seen with 100 pg of DNA input. Also, Meiklejohn and Robertson (2017) found that in their evaluation experiment, robust and reproducible profiles can be collected with ≥ 0.2 ng of neat DNA or forensic samples. The panel herein was applied to real casework samples and the results collected showed that the full profile was seen in 27 out of the 60 samples. Full profiles were observed down to 0.06 ng input DNA. The results of the sensitivity study that was conducted in the previous chapter, full SNP profiles seen down to 0.25 ng DNA input using the 21 PCR cycles. DNA input of 0.05 ng also led to complete profiles in two replicates. Partial profiles with 11, 5 and 14 dropouts were observed in 0.01 ng replicates and one SNP left in the third replicate of 0.05 ng. However, when the same samples were profiled with the GlobalFiler STR kit the percentage profile recovered was in most cases as high or higher than with the SNPs. This can be explained to some degree by the samples collected not being highly degraded, but also that the STR chemistry is more robust. This is illustrated in the samples that gave full STR profiles, but no SNP profile, which is presumably due to PCR inhibition.

The coverage values of autosomal and Y SNPs were reviewed and compared with the coverages collected in the previous experiment, which was applied to population samples (see Chapter 3). Between the two series of experiments of the Identity panel, the autosomal SNPs rs214955 (100 bp) and rs722290 (173 bp) received the lowest

coverage values in 10 and 9 samples, respectively. In contrast, in the previous experiment, SNP rs214955 showed the lowest coverage in 58 samples. For Y SNPs, SNP M479 provided low coverage across multiple samples in both studies. The lowest coverage out of all SNPs was seen for M479 in 65 population samples and 25 casework samples. In addition to M479, it was found that rs2032599 showed lower coverages in 12 samples; this is similar to data reported in two previous studies (Eduardoff et al., 2016; Ochiai et al., 2016).

The available MPS panels for forensic investigations can provide useful information together with the primary genetic marker STR used in forensic DNA typing also may play an important role with DNA databanks in forensic laboratories when there is no match found. The presence of such information may be important in some complex or cold cases and also assist in the ongoing investigation of the case in predicting biogeographical ancestry or typing highly degraded samples (Budowle et al., 2017; Hollard et al., 2017).

Furthermore, the haplo-groups predictions based on Y-SNPs can add valuable data to the forensic scientists and the investigators through paternal biogeographical ancestry inference (Ochiai et al., 2016). Through analysis of 34 Y-SNPs included in the Precision ID Identity panel, a total of 8 different Y-haplo-groups were observed among the samples analyzed, two haplo-groups were not predicted in the population experiment which are H1 (M69) and IJ (M429). The Clade H defined by M69 is widespread in many Indian populations (Debnath et al., 2011). The report generated from the HID SNP Genotyper provides for each predicted haplo-group, a description that showed the haplogroup evolution and lineage distribution. However, for the haplo-group IJ there is no explanation regarding its origin. Karafet et al. (2008) described major changes in the topology of the parsimony tree and provided names for new and rearranged lineages within the tree following the rules presented by the Y Chromosome Consortium in 2002. They reported that there were seven mutations (P123, P124, P125, M429, P126, P127, P129, and P130) that joined haplo-groups I and J into the "IJ" clade. The bifurcation of IJ-M429 occurred roughly about forty-one thousand years ago. IM170 probably evolved from IJ-M429 in South Eastern Europe, and JM304 evolved from IJ-M429 in South East Asia (Clair, 2020). Thermo Fisher states that to successfully identify

a haplo-group, at least 30% of the Y SNPs should have a valid genotype (Thermo Fisher Scientific, 2017b). There were 17 samples which showed "No haplogroup found", although one was a female and the remaining samples were partial profiles. In addition, it was shown in one sample that had a drop out at locus AMEL in Y allele (S28-BC28-P5). However, a haplo-group IJ observed in a sample (S29-BC29-P5) that had a similar issue of Y allele drop out (STR results) and only 5 Y SNPs dropped. Similarly, IJ was the predicted haplogroup for two samples (S3-BC3-P5,) and (S4-BC4-P5) with 5 YSNPs missed the genotypes (rs9786139 (120 bp), rs16981290 (161 bp), rs17250845 (94 bp), L298 (97 bp), P256 (119 bp)).

Y chromosome haplo-groups determined by the combination of allelic states at binary SNP loci show clearly different geographic distributions and may predict the ancestral and geographic origins of unknown casework samples (Muro et al., 2011). As mentioned in the previous chapter, the results obtained from the population experiments of analyzing 34-SNPs, included in the Precision ID Identity panel, have been observed to some extent that the Qatari population of people share the same genetic patterns with neighboring regions.

4.5 Conclusion

In conclusion, the recent targeted MPS technologies are highly promising because of their large multiplex and with inclusion of Y-SNPs to study the male lineage. The results presented in this chapter from this preliminary study showed that the Precision ID Identity panel may be part of forensic casework analysis for human identification study employing ancestry parental lineage with the STR standard tool. In turn, this may be useful in some cold cases as well as for some samples that may show some issues during the analysis and in complex kinship cases with mutations.

However, the overall success of the SNP profiling was lower than that seen for standard STR testing, which can be explained by two factors. Firstly, the current chemistry only allows 6 µl of template DNA to be added, in comparison the STR kits allow up to 15 µl of template. Secondly, where sufficient DNA is available the chemistry is not so robust to inhibition. This work is based on a range of 60 casework samples, but the limitations of the SNP system are evident.

CHAPTER: 5 Evaluation of Precision ID Ancestry Panel in Qatar population by next generation sequencing using the Ion Torrent Personal Genome Machine (PGM)

Part one: Sequencing performance, allele frequencies and other forensic parameters of the panel in the Qatari population.

5.1 Introduction

Single nucleotide typing with massively parallel sequencing (MPS) platforms and bespoke forensic panels can potentially play an important role in forensic laboratories and can be considered as a complementary tool to STR typing in forensic casework analysis. This can be related to identity as discussed in the previous chapter, but also ancestry and phenotype. Several sets of Ancestry Informative Markers (AIMs) have been identified comprised of markers that demonstrate relatively large differences in allele frequencies between different biogeographical groups (Pardo-Seco et al., 2014).

The Precision ID Ancestry panel (TFS) was one of the first panels to be released for forensic studies. The Ancestry Informative Single nucleotide polymorphisms (AISNPS) panel comprises 165 markers. The panel includes 55 from the Kidd panel (Kidd et al., 2014) with an average amplicon size 130 bp and 123 SNPs from the Seldin panel (Kosoy et al., 2009) with an average amplicon size of 120 bp (TFS). The Manufacturer stated that the kit would amplify 1 ng of template DNA (optimally, although it would work with lower amounts) followed by MPS on the Ion Torrent Personal Genome machine (PGM) and/or Ion S5 (TFS).

Two forms of the ancestral predictions could be performed with the use of plug-in HID SNP Genotyper which were Admixture Prediction and Population Likelihood estimates. The Admixture Prediction provided an estimate of the overall ancestry contribution in each sample. This is based on seven root populations (Africa, America, East Asia, Europe, Oceania, South Asia, and Southwest Asia) and Population Likelihood that calculates the most likely population of origin of the DNA profile (TFS).

The main aim of this chapter of the study was to evaluate the HID-Ion Ampliseq Ancestry Panel within the Qatar population, optimising the workflow and to study the sensitivity of the panel and its potential value in defining the source population of forensic evidence.

Objectives

- a) To evaluate massively parallel sequencing workflow within the Forensic Laboratory of Qatar.
- b) To collect and genotype 300 unrelated Qatari individuals from the eight municipalities that makeup the State of Qatar.
- c) To evaluate the panel sensitivity using 5 different DNA concentration (1 ng, 0.5 ng, 0.25 ng, 0.05 ng and 0.01 ng).
- d) To evaluate the forensic statistical parameters generated from the 300 Qatari individuals and the forensic parameters for the 165 AISNPs.
- e) To investigate ancestry inference ability of the Precision ID Ancestry Panel in Qatar population
- f) To evaluate the population structure using STRUCTURE for the eight municipalities studied.

5.2 Materials and Methods

The materials and methods relevant to this Chapter were described in Chapter 2.

5.3 Results

5.3.1 Sensitivity study

The sensitivity of the Precision ID Ancestry Panel was tested by using a buccal swab sample from male volunteer. The samples were quantified for DNA with real-time PCR and then serially diluted to provide five template amounts: 1 ng, 0.5 ng, 0.25 ng, 0.05

ng and 0.01 ng. The dilutions were prepared in triplicate for library preparation with 21 PCR cycles. Different barcodes were used for the libraries and they were sequenced on the same Ion 316 Chip.

5.3.1.1 Sensitivity study results

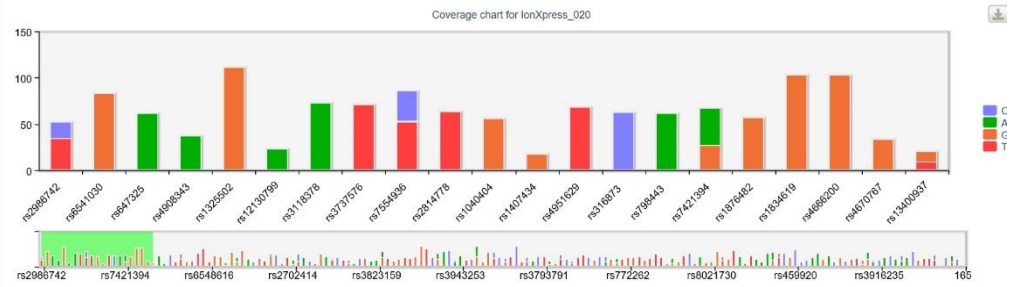
From HID SNP Genotyper the sensitivity results were reviewed as a sample table, which included the genotype list for all the samples across all hotspots. Also, allele coverage was generated for the experiments as tables and charts. The charts tab showed the coverage across the hotspots whereas the table tab shows the allele details such as genotype, coverage and quality check information (QC check). The sample genotypes flagged according to the HID SNP Genotyper QC filter, which included coverage, percent positive coverage, major allele frequency and the genotype as explained in Table 3.1 (Chapter 3).

The table contains the samples which include the genotypes and allele coverage for the experiment. The data were exported from the plugin and an excel file which were prepared including the sample name, their barcodes, the genotype, the number of loci that have drop out and the confidence level. Most loci dropped out in the 0.01 ng replicates, partial profiles were seen with the 0.05 ng replicates (Table 5.1); (Figure 5.1) and (Figure 5.2).

Table 5.1. Data showing the results of the sensitivity study. Confidence level is with reference to the ancestry prediction.

Samples	DNA concentrations	#of loci drop out	Confidence levels
Replicate 1	1 ng	0	High
	0.5 ng	0	High
	0.25 ng	0	High
	0.05 ng	6	High
	0.01 ng	No Data	Low
Replicate 2	1 ng	0	High
	0.5 ng	0	High
	0.25 ng	0	High
	0.05 ng	2	High
	0.01 ng	27	Low
Replicate 3	1 ng	0	High
	0.5 ng	0	High
	0.25 ng	0	High
	0.05 ng	0	High
	0.01 ng	19	High

(A)



(B)

Chrom	Positio...	HotSpot ID	QC	Cov	A Reads	C Reads	G Reads	T Reads	Pos Cov	Neg Cov	Pos Cov...	Genoty...	GQ	Maj. Allele Fre...
chr1	159174...	rs2814778	█	101	0	45	0	56	50	51	49.5	CT	99	55.45
chr1	168159...	rs1040404	█	73	36	0	37	0	36	37	49.32	AG	99	50.68
chr1	186149...	rs1407434	█	38	8	0	30	0	23	15	60.53	AG	16	78.95
chr1	212786...	rs4951629	█	149	0	0	0	149	82	67	55.03	TT	67	100
chr1	242342...	rs318873	█	90	0	45	0	45	38	52	42.22	CT	99	50
chr2	7988275	rs798443	█	101	100	1	0	0	40	61	39.6	AA	45	99.01
chr2	14756349	rs7421394	█	115	115	0	0	0	62	53	53.91	AA	52	100
chr2	17362568	rs1878482	█	101	0	0	101	0	55	46	54.46	GG	46	100
chr2	17901485	rs1834819	█	94	0	0	94	0	51	43	54.26	GG	42	100
chr2	29539411	rs4666200	█	136	135	0	1	0	72	64	52.94	AA	53	99.28
chr2	37941396	rs4670767	█	68	0	0	68	0	39	29	57.35	GG	31	100
chr2	79864923	rs13400937	█	32	0	0	16	16	14	18	43.75	GT	73	50
chr2	109513...	rs3827780	█	101	99	0	2	0	55	46	54.46	AA	30	98.02
chr2	109579...	rs280690	█	28	11	17	0	0	15	13	53.57	AC	43	60.71
chr2	136707...	rs6754311	█	92	0	92	0	0	50	42	54.35	CC	41	100
chr2	145769...	rs10496971	█	115	0	0	0	115	53	62	46.09	TT	52	100
chr2	158887...	rs10497191	█	95	0	95	0	0	45	50	47.37	CC	43	100

Figure 5.1. The above diagram and table illustrating the coverage bar chart (A) and coverage table (B) for sample 0.01 ng, replicate 2, with 23 loci dropped out. The example in B shows rs1407434 failing the Quality Control; major allele frequency balance is 78.95 which is >65% (MAF).

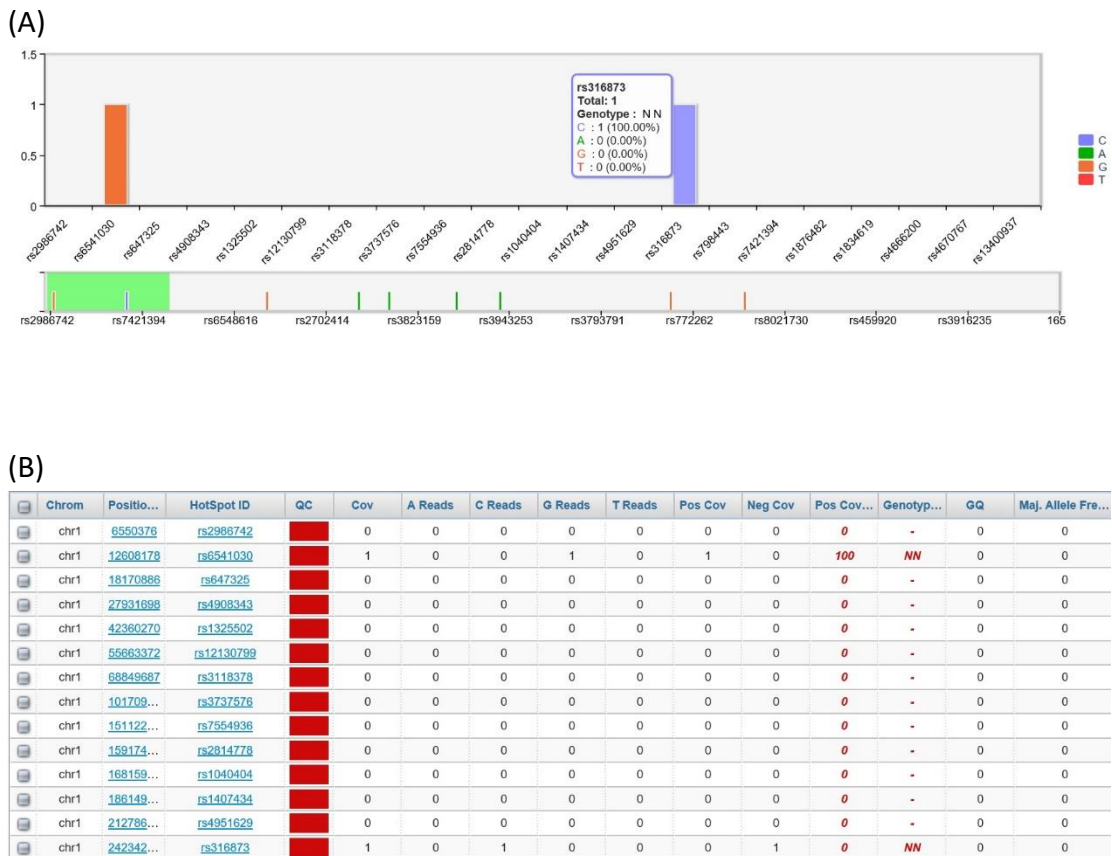


Figure 5.2. The above bar chart diagram (A) and (B) table illustrating the coverage charts and coverage tables, respectively for sample 0.01 ng, replicate 1, with no profile.

5.3.2 Quality of the sequencing runs

In this study, a total of 300 unrelated Qatari individuals were collected from the eight municipalities that makeup the State of Qatar.

The samples were amplified according to the manufacturer’s protocol using 1 ng of DNA at 21 cycles. Library preparation was performed using the Ion AmpliSeq™ Library Kit. Template preparation and chip loading were performed on the Ion Chef™ System. The Sequencing was carried out on the Ion PGM™ instrument (TFS) on an Ion 316 Chip v2 (TFS). A total of 10 chips were used (30 samples per chip).

Each run was evaluated, and the quality was subsequently, expressed for each chip in a run summary exported as a pdf file. The upper part of the report page consists of three rectangular sections showing the quality of the run (Figure 5.3).

Run Summary

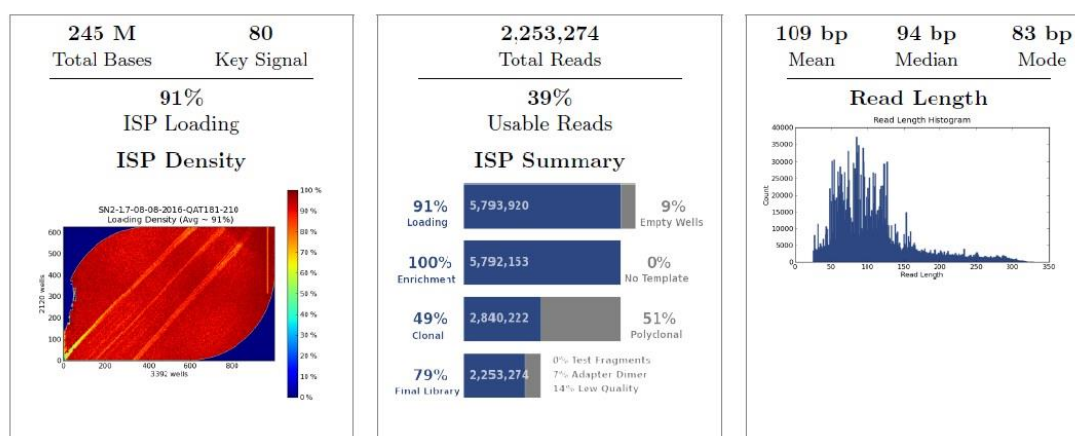


Figure 5.3. Diagrams showing the three sections that evaluate the quality of the run (run summary).

The first section shows the average Ion Sphere™ Particles (ISPs) density expressed by a heat map; red indicated a good loading, yellow heat map indicates less ISP, but it is acceptable while blue and green indicate poor loading.

With the ten chips sequenced the highest average ISP density was 91%, achieved in two runs (Figure 5.4) and the next best 7 runs ranged from 83% to 66% (Table 5.2). The lowest average ISP density was seen in one run with 46% (Figure 5.5). The second part of the report showed the ISP summary. The loading percentage always corresponded to the average loading density (ISP density %) which was the number of ISPs in the wells. The percentage of wells that included ISPs while the empty wells indicated the total number of the wells without ISPs, taking into account that the 316 chip contained 6 million wells. The enrichment percentage provided information on how many of ISPs held DNA on them and the quality of ISP (i.e. clonal ISP holding one template) or polyclonal (ISP holding two or more templates), Finally, the percentage of the final library indicated how many reads were in the final library and recorded in the output. The third section of the run report shows a histogram of the read length and it also describes the distribution of the read length of HID Ancestry panel. The mean read length for all the ten chips were: 101 bp, 101 bp, 100 bp, 97 bp, 106 bp, 106 bp, 109 bp, 106 bp, 99 bp and 96 bp (Chip 1-10 in order).

Table 5.2. Data showing the Average ISP density for each chip used. The lowest value is highlighted.

Chips#	Average ISP Density
Chip #1 (QAT1-30)	83%
Chip #2 (QAT31-QAT60)	80%
Chip #3 (QAT61-QAT90)	80%
Chip #4 (QAT91-QAT120)	80%
Chip #5 (QAT121-QAT150)	66%
Chip #6 (QAT151-QAT180)	46%
Chip #7 (QAT181-QAT210)	91%
Chip #8 (QAT211-QAT240)	78%
Chip #9 (QAT241-QAT270)	91%
Chip #10 (QAT271-QAT300)	79%

(A)

(B)

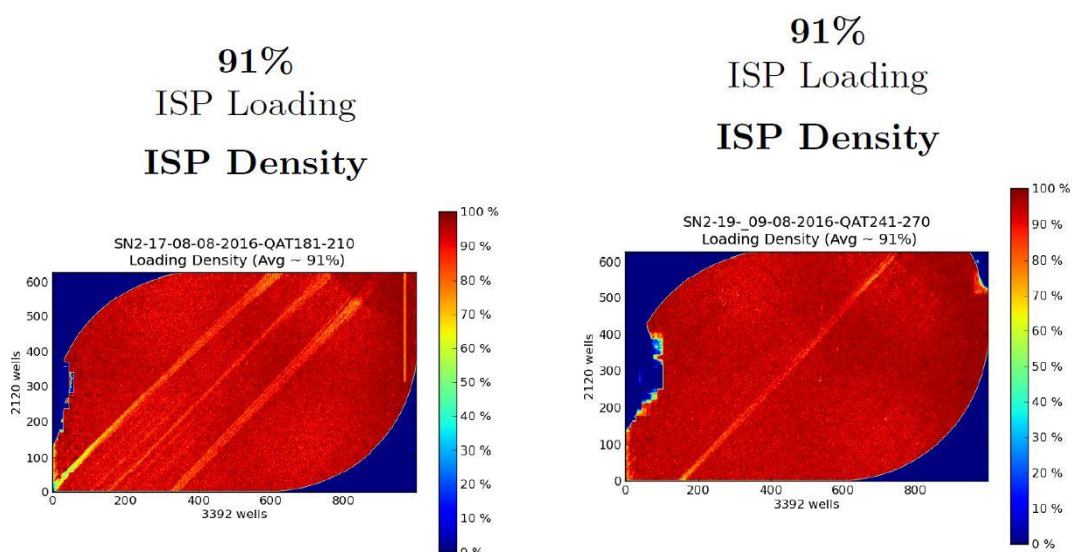


Figure 5.4 . Images of sequencing chips with highest ISP loading density. (A) Chip#7 QAT181-QAT210). (B) Chip#9 (QAT241-QAT270).

46%

ISP Loading

ISP Density

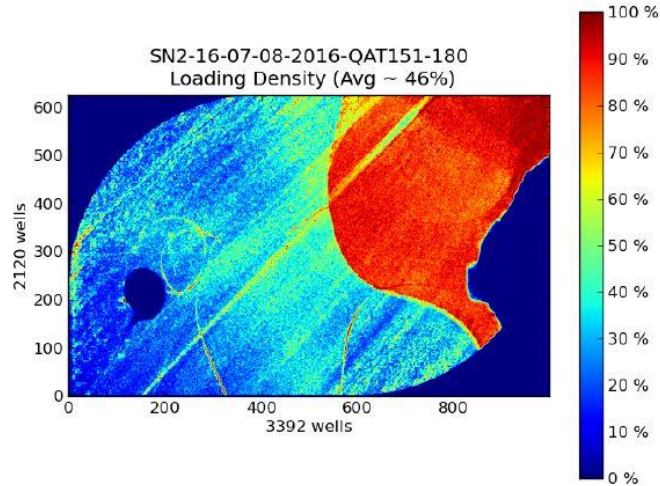


Figure 5.5. An image of chip#6 (QAT151-QAT180) with the lowest average ISP density.

5.3.3 Sequence Coverage

Coverage or sequencing depth represents the number of aligned reads that cover a hotspot. The coverage results were reviewed in terms of coverage table and coverage chart. From the plug-in HID SNP Genotyper the allele coverage pane results were generated for the 10 chips as tables and charts. The charts showed the coverage across the hotspot whereas the table tab shows the coverage allele details such as genotype, coverage and quality check information (QC check). The results were exported for each chip in separate Excel sheet (Figures 5.6 and 5.7) and Table 5.3.

Table 5.3. Data showing the average mean depth obtained from 10 chips. The highlighted row represents the Chip with the lowest coverage.

Chip#	Chip loading%	Total Reads	Averages mean depth
Chip #1	83%	2,559,400	310.07
Chip #2	80%	2,211,942	260.89
Chip #3	80%	2,409,064	311.155
Chip #4	80%	2,506,029	336.741
Chip #5	66%	2,771,415	286.79
Chip #6	46%	1,605,299	158.29
Chip #7	91%	2,253,274	263.903
Chip #8	78%	1,451,762	216.304
Chip #9	91%	2,451,762	336.809
Chip #10	79%	1,945,627	258.38

(A)

Chrom	Positio...	HotSpot ID	QC	Cov	A Reads	C Reads	G Reads	T Reads	Pos Cov	Neg Cov	Pos Cov...	Genoty...	GQ	Maj. Allele Fre...
chr1	6550376	rs2986742	■	59	0	0	0	59	29	30	49.15	TT	27	100
chr1	12608178	rs6541030	■	65	0	0	65	0	24	41	36.92	GG	29	100
chr1	16170886	rs647325	■	82	82	0	0	0	45	37	54.88	AA	37	100
chr1	27931698	rs4908343	■	70	70	0	0	0	35	35	50	AA	32	100
chr1	42360270	rs1325502	■	65	34	0	31	0	37	28	56.92	AG	99	52.31
chr1	55663372	rs12130799	■	59	59	0	0	0	31	28	52.54	AA	27	100
chr1	68849687	rs3118378	■	53	53	0	0	0	22	31	41.51	AA	24	100
chr1	101709...	rs3737576	■	84	0	0	0	84	42	42	50	TT	38	100
chr1	151122...	rs7554936	■	61	0	28	0	33	38	23	62.3	CT	99	54.1
chr1	159174...	rs2814778	■	74	0	74	0	0	32	42	43.24	CC	33	100
chr1	168159...	rs1040404	■	52	52	0	0	0	27	25	51.92	AA	23	100
chr1	186149...	rs1407434	■	42	0	0	42	0	23	19	54.76	GG	19	100
chr1	212786...	rs4951629	■	92	0	1	0	91	55	37	59.78	TT	33	98.91
chr1	242342...	rs316873	■	72	0	72	0	0	33	39	45.83	CC	32	100
chr2	7968275	rs798443	■	55	55	0	0	0	24	31	43.64	AA	25	100
chr2	14756349	rs7421394	■	72	43	0	29	0	41	31	56.94	AG	99	59.72
chr2	17362568	rs1876482	■	69	0	0	69	0	41	28	59.42	GG	31	100
chr2	17901485	rs1834619	■	56	0	0	56	0	24	32	42.86	GG	25	100
chr2	29638411	rs4666200	■	102	53	0	49	0	52	50	50.98	AG	99	51.96

(B)



Figure 5.6. Data showing the Coverage results for sample QAT271 with no drop out.

(A) Coverage table and (B) Coverage bar charts.

(A)

Chrom	Positio...	HotSpot ID	QC	Cov	A Reads	C Reads	G Reads	T Reads	Pos Cov	Neg Cov	Pos Cov...	Genoty...	GQ	Maj. Allele Fre
chr1	6550376	rs2986742	■	0	0	0	0	0	0	0	0	-	0	0
chr1	12606178	rs6541030	■	0	0	0	0	0	0	0	0	-	0	0
chr1	18170886	rs647325	■	0	0	0	0	0	0	0	0	-	0	0
chr1	27931698	rs4908343	■	3	3	0	0	0	1	2	33.33	NN	0	0
chr1	42360270	rs1325502	■	0	0	0	0	0	0	0	0	-	0	0
chr1	55663372	rs12130799	■	2	2	0	0	0	1	1	50	NN	0	0
chr1	68849687	rs3118378	■	0	0	0	0	0	0	0	0	-	0	0
chr1	101709...	rs3737576	■	0	0	0	0	0	0	0	0	-	0	0
chr1	151122...	rs7554936	■	0	0	0	0	0	0	0	0	-	0	0
chr1	159174...	rs2814778	■	1	0	0	0	1	0	1	0	NN	0	0
chr1	168159...	rs1040404	■	0	0	0	0	0	0	0	0	-	0	0
chr1	186149...	rs1407434	■	0	0	0	0	0	0	0	0	-	0	0
chr1	212786...	rs4951629	■	0	0	0	0	0	0	0	0	-	0	0
chr1	242342...	rs316873	■	1	0	1	0	0	0	1	0	NN	0	0
chr2	7968275	rs798443	■	2	2	0	0	0	1	1	50	NN	0	0
chr2	14756349	rs7421394	■	4	1	0	3	0	2	2	50	NN	0	0
chr2	17362668	rs1876482	■	2	0	0	2	0	0	2	0	NN	0	0

(B)

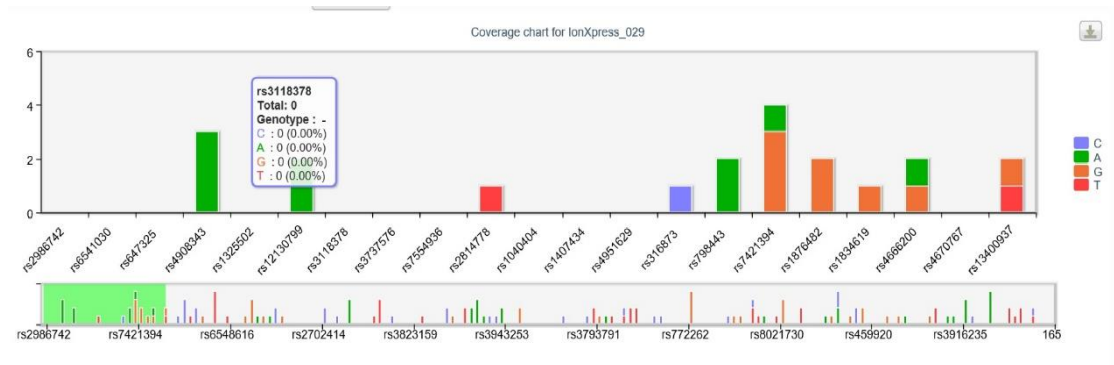


Figure 5.7. Data showing the Coverage results for sample QAT299, which had no genotype. (A) Coverage table and. (B) Coverage bar charts.

The default setting for the coverage was minimum coverage of 6, which was used for the analysis. The highest coverage was seen at rs3907047 (3675). Coverage values below 6 were seen at rs13400937, rs1950993, rs260690, rs2306040, rs11652805, rs6990312, rs1407434, rs6990312, rs1296819, rs310644. An Excel file was prepared for the coverage for all the 299 Qatari samples (one had no genotype) with the 165 Ancestry SNPs, showing the maximum, minimum and the average coverage values. Also, the highest and lowest coverage value among all the samples (Table 5.4 and 5.5).

Table 5.4. Data showing the SNPs with the highest coverage among the 299 samples.

SNP/ID	#samples	SNP/ID	#samples
rs4955316	134	rs459920	2
rs1837606	97	rs385194	1
rs200354	17	rs12544346	1
rs735480	11	rs10513300	1
rs3907047	9	rs9530435	1
rs316873	6	rs1325502	1
rs3793451	5	rs6556352	1
rs316598	4	rs17642714	1
rs2196051	3	rs10839880	1
rs1079597	2	rs4666200	1
Total= 299			

Table 5.5. Data showing the SNPs with the lowest coverage among the 299 samples.

SNP/ID	#samples	SNP/ID	#samples
rs13400937	73	rs1871534	3
rs1950993	65	rs310644	3
rs260690	54	rs7997709	3
rs6990312	23	rs10007810	3
rs1296819	22	rs2306040	2
rs772262	9	rs12629908	2
rs1407434	8	rs2966849	1
rs4833103	5	rs3916235	1
rs4670767	4	rs11652805	1
rs1040045	4	rs1760921	1
rs192655	4	rs2986742	1
rs4471745	3	rs2196051	1
rs10512572	3		
Total = 299			

5.3.4 Performance of the Precision ID Ancestry Panel

A total of 300 samples were typed and only one sample had no genotype. All the genotype results for all the samples were exported as excel files chip-by-chip. Partial profiles were seen in 59 samples from the 300 samples (Table 5.6).

Table 5.6. The data showing the overall genotype for the 300 samples in each chip with chip loading%.

Chips#	Samples#	Genotypes			Chip loading%
		Full	Partial	No	
Chip#1	QAT1-QAT30	30	-	-	83%
Chip#2	QAT31-QAT60	30	-	-	80%
Chip#3	QAT61-QAT90	29	1	-	80%
QAT#4	QAT91-QAT120	28	2	-	80%
Chip#5	QAT121-QAT150	16	14	-	66%
Chip#6	QAT151-QAT180	14	16	-	46%
Chip#7	QAT181-QAT210	26	4	-	91%
Chip#8	QAT211-QAT240	30	-	-	78%
Chip#9	QAT241-QAT270	24	6	-	91%
Chip#10	QAT271-QAT300	23	6	1	79%

5.3.5 Statistical analysis of SNPs data

All the Excel files including the genotypes for the 300 Qatari samples were exported from the Torrent Server and were used in statistical calculations. PowerStats V12 software Microsoft Excel Workbook template (Tereba, 1999) was used for allele frequencies, power of discrimination (PD), random match probability (MP), typical paternity index (TPI), probability of exclusion (PE) and polymorphism information content (PIC) calculations. The Exact test for Hardy –Weinberg equilibrium (p -value), observed heterozygosity (H_o), expected heterozygosity (H_e) were calculated using Arlequin (Excoffier et al. 2010).

5.3.5.1 Allele Frequencies of 165 HID Ancestry SNPs in Qatar Population

The allele frequencies of the 165 SNPs were calculated and are listed in (See Figure 5.8 and Appendix 11).

5.3.5.2 Hardy-Weinberg Equilibrium (HWE)

The p - values observed (Obs.Het) and expected heterozygosity (Exp.Het) values for Hardy-Weinberg equilibrium were performed and summarized (See Figure 5.9 and Appendix 12).

Ten SNPs (rs2814778, rs3823159, rs705308, rs1871534, rs3814134, rs200354, rs735480, rs1426654, rs11652805, rs7251928) showed deviation from HWE at $p < 0.05$. After applying Bonferroni correction 0.0003 (0.05/165) all the loci were in HWE. Three SNPs rs3811801, rs671 and rs1800414 were monomorphic, which is not unexpected for ancestry SNPs. The observed heterozygosity values were in the range of 0.01087 (rs2042762) to 0.55442 (rs9845457) whereas the expected heterozygosity values ranged from 0.01083 (rs2042762) to 0.50082 (rs459920).

5.3.5.3 Forensic Parameters of 165 SNPs

The Combined Probability of Match (CPM) of the 165 SNPs was 1.35×10^{-42} , the Combined Power of Discrimination 0.99999 and the Combined Power of Exclusion 0.999998. With this panel the SNP showed the highest discrimination power (PD) was rs4908343 (0.640) and the SNP with the highest power of exclusion was rs2814778 (0.375) (See Figure 5.10 and Appendix 13).

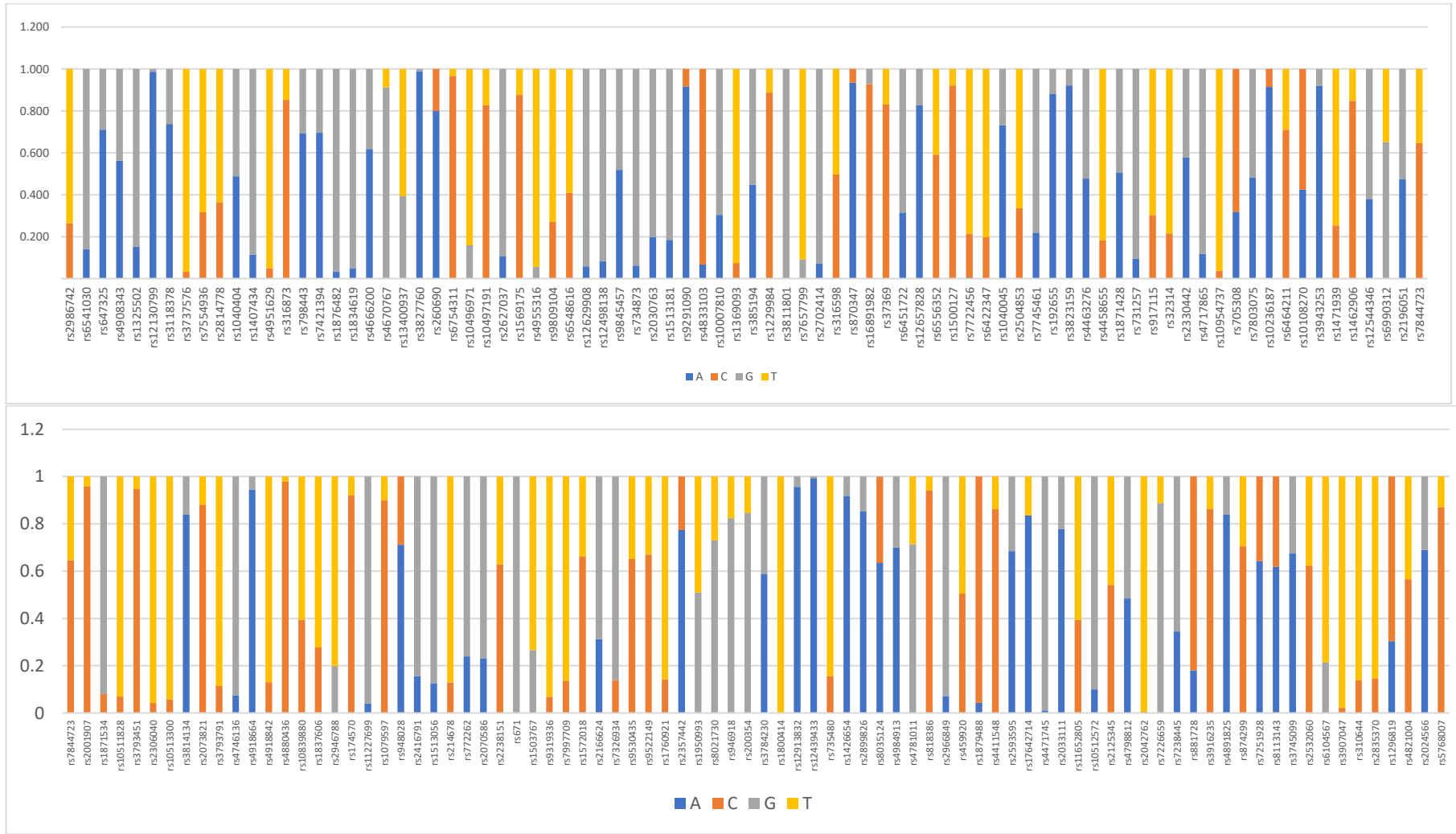


Figure 5.8. Column plot of allele frequencies of 165 autosomal SNPs of Precision ID Ancestry panel in Qatar population.

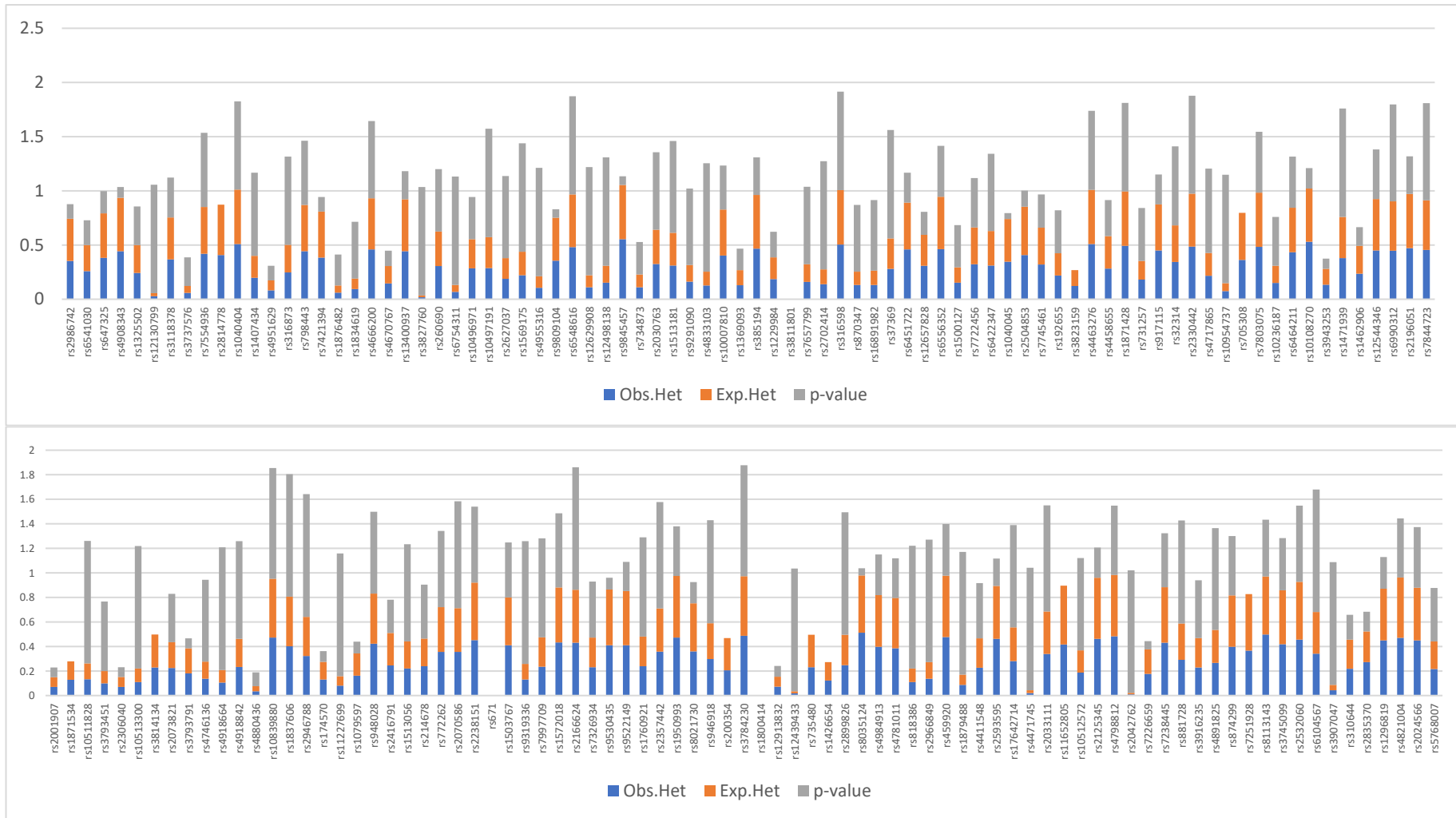


Figure 5.9. Bar plots of Observed heterozygosities (Obs.Het) expected heterozygosities (Exp.Het) and p- values across 165 SNPs of the Precision Ancestry panel. Monomorphic SNPs are: rs3811801, rs671 and rs1800414.

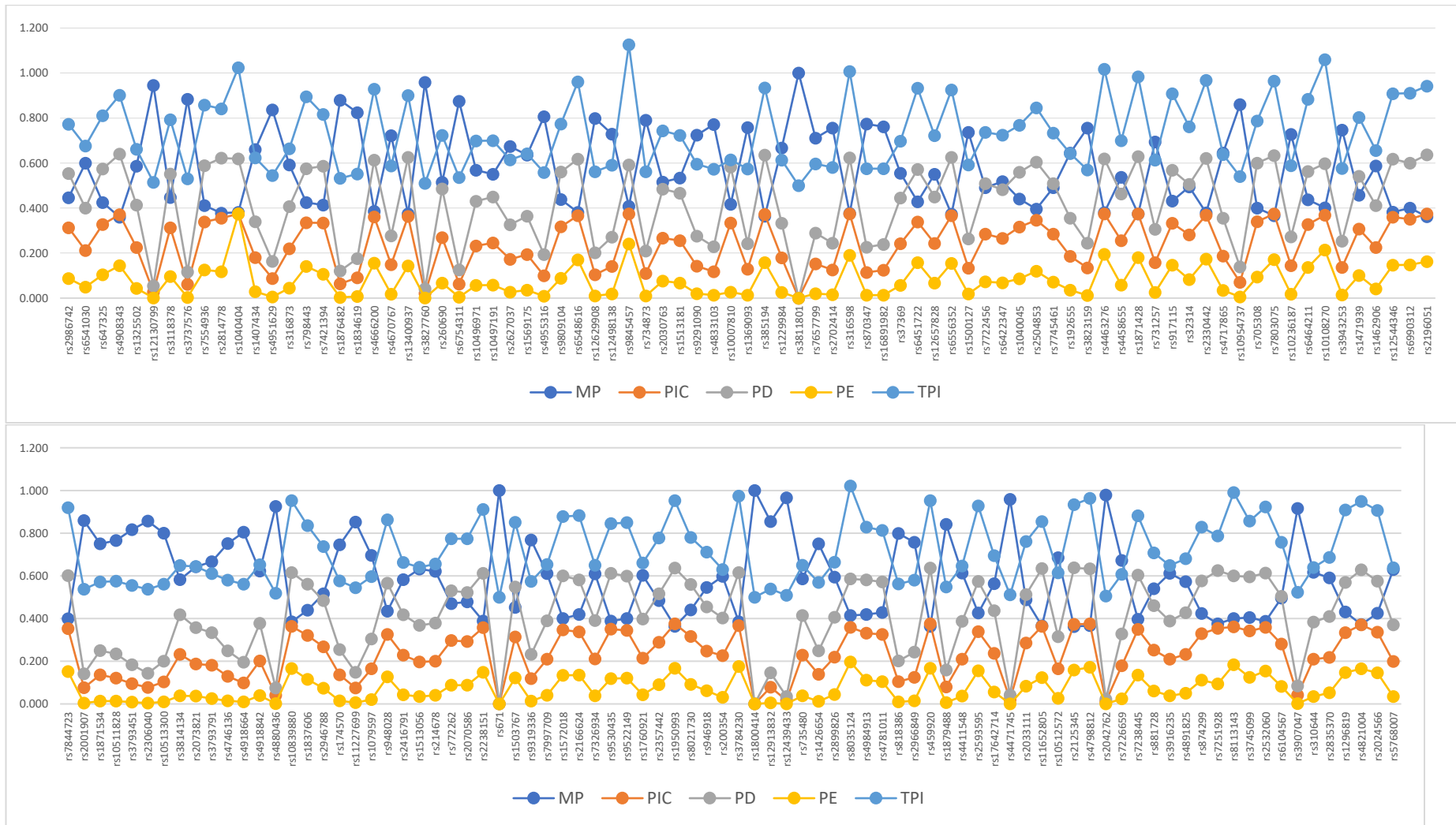


Figure 5.10. Plot graph showing the forensic parameters of 165 SNPs of The Precision ID Ancestry panel (PM: probability of match, PD: power of discrimination, PE: Power of Exclusion, PIC: polymorphism information content, and TPI: Typical paternity index).

5.4 Discussion

In this part one of the study, 300 unrelated Qatari individuals from the eight municipalities were first typed to investigate the forensic parameters and ancestry inference ability of the Precision ID Ancestry Panel in Qatar population and the Ion PGM.

The sensitivity experiment showed that the alleles dropout was seen when the samples had an input of DNA 0.05 ng and 0.01 ng. One of the three replicates of 0.05 ng gave a full genotype, 6 and 2 loci drop-out were seen in the other two replicates. The results obtained from 0.01 ng three replicates were no data, 27 loci and 19 loci drop-out. A template target of 1 ng was used for reference samples, however 59 samples suffered from some allele drop out in different SNPs.

The samples passed through a series of steps before the sequencing starting from the from preparation of the template to construct the libraries. It is possible that dilution step may affect the sample in having either low DNA template or low library and that may lead to partial genotype.

The Ion Chef™ System was used for the template preparation of pooled Ion Ampliseq libraries and chip loading. Ion Chef™ provides an automation system for template preparation including emulsion PCR, enrichment and loading two chips for sequencing on Ion PGM. The Ion Chef™ enables time and labour savings, with less than 15 minutes of hands-on time for setup. In a study by Mogensen et al., (2015), they found that the workload and the number of manual pipetting steps were decreased while the chip loading efficiency increased using the Ion Chef™ System. In their study, the number of pipetting steps was reduced from an average of 70 for ten libraries to two with the Ion Chef™. The hands-on time was reduced from 1 hour to 15 min. Consistent chip loading helped to get very good loading values. Low values of ISP density impact on the results can lead to partial genotypes and allele drop out (Mogensen et al., 2015).

The summary run report from the Torrent browser was used to assess the quality of the Ion Torrent PGM run through the statistics and quality metrics for the run. These included the Ion Sphere™ Particle (ISP) density. Chip #6 had the lowest average ISP density (46%), and the high number of samples suffered from drop out was seen in Chip #6. Two reasons may cause this issue either the liquid was not removed after the

chip check or too much liquid remained in the chip after loading (Thermo Fisher Scientific, 2015). This is one of the most important elements which has direct relation and effect on the sensitivity and genotyping accuracy of NGS systems applied to forensic SNP typing. With Ion PGM, the number of wells per chip that can be filled with ISPs defines the number of possible reads. Sample pooling, template preparation influencing the number of non-templated and polyclonal ISPs and loading efficiency can influence the final number of successfully read ISPs (monoclonal reads) (Eduardoff et al., 2015). Coverage value (depth of coverage/sequencing depth) refers to the number of times each base has been read in the sequencing run. As shown above in Table 5.5, the list of SNPs showed the lowest coverage across the samples among them four loci rs13400937, rs1950993, rs260690 and rs6990312 that showed poor coverage. In twenty-two samples SNP rs1296819 had the lowest coverage; this SNP also performed poorly in other studies (Al-Asfi et al., 2018; Pereira et al., 2017; Santangelo et al., 2017).

Therefore, it is necessary to optimize the procedure of HID Ion Ampliseq Ancestry Panel on forensic evidence, including samples that have different degrees of degradation or samples with low DNA. This scope includes amplifying poor-quality samples with varying PCR cycle numbers to determine the optimum number of PCR cycles that will be suitable for sub-optimum samples.

This study detected the 3 SNPs rs3811801 (GG), rs671(GG) and rs1800414(TT) were monomorphic in the Qatari population. It has also been reported that these three SNPs were monomorphic in the following studies in the three main ethnic groups in Ecuadorian population (Santangelo et al., 2017), Basques population (García et al., 2017), Danes and Somali populations (Pereira et al., 2017) and in a Greenlander population (Espregueira-Themudo et al., 2016). The ideal AIM is monomorphic in one population and highly abundant in another (Rosenberg et al., 2003). Those monomorphic SNPs in Qatari population may be highly informative if these SNPs are polymorphic, especially if the 'non-Qatari' allele is present at a high frequency. These three SNPs were not recorded as monomorphic SNPs in the study which was applied to seven Asian populations (Lee et al., 2018) as well as the Chinese individuals from Tibeto-Burman study (Wang, Z. et al., 2018). However, the Ensembl genome browser

shows that these three markers are fixed in all the European, Native Amerindian (American Indian) and African samples reported to date, but they are polymorphic in East Asian populations (Santangelo et al., 2017). As such, they have value in differentiating East Asian populations from others.

While it is not the primary intention, the Precision ID Ancestry panel can be used as a tool in forensic DNA testing for identification. In this panel the SNP rs9408343 had the highest discrimination power (PD) (0.640). The Combined Power of Discrimination (CPD) (0.99999) and the Combined Probability of Match (CPM) was 1.35×10^{-42} allowing a very good level of discrimination in forensic cases. SNP rs2814778 had the highest Power of Exclusion (PE) (0.375) and the combined Power of Exclusion was (0.999998). These values were then compared with those of published populations (Table 5.7).

Table 5.7. Data showing the comparison of 165 autosomal SNPs in the Precision ID Ancestry Panel data between Qatar population and other world populations.

Populations	CPM	CPE	References
Qatar	1.35×10^{-42}	0.999998	-
Basques	3.13×10^{-35}	0.999972	(García et al., 2017)
Liangshan Yi	3.520×10^{-46}	0.999999984	(Wang, Z. et al., 2018)
Chengdu Tibetan	1.150×10^{-46}	0.999999978	(Wang, Z. et al., 2018)
Liangshan Tibetan	1.783×10^{-45}	0.999999966	(Wang, Z. et al., 2018)
Qinghai Tibetan	3.306×10^{-49}	0.999999999	(Wang, Z. et al., 2018)

In conclusion, it is possible that from the results mentioned above in this part of the study and also based on their strong statistical values, the panel can be a useful forensic tool, especially when combined with the STRs for forensic investigation. The data can also help in the investigation of forensic and paternity cases, especially in those with mutations and/or in complex cases with relatives involved.

Part two: Ancestry Inference of the Precision ID Ancestry panel and STRUCTURE analysis in Qatar population

5.5 Results

5.5.1 Ancestry prediction with the default HID SNP Genotyper

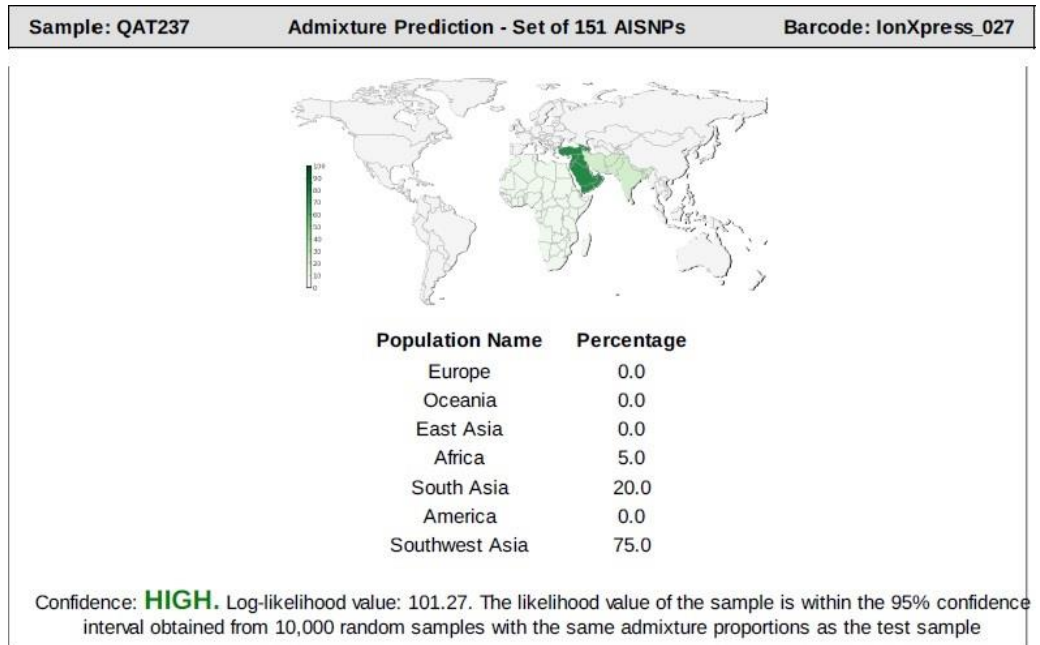
The Plug-in; HID SNP Genotyper version 4.3.2 generate a pdf file includes the two forms of ancestry prediction as admixture prediction and population likelihood results. The Admixture prediction finds the best ancestral composition of a sample. From the seven root populations included in the plug-in: Africa, America, Southwest Europe (Middle East), Europe, Oceania, East Asia and South Asia, the admixture prediction finds the most likely combination of these population that best explains the sample genotype. The data set consists of 151 SNPs Ancestry markers (AIMs) which are used to estimate the admixture prediction calculations of the seven populations according to their genotype frequencies. The software reference only 151 of the 165 markers studied based on population data available to Thermo Fisher Scientific (TFS), which are a subset of Kidd's 55 AIMs and Seldin's 128 Aims.

The Plug-in also shows the population likelihoods values. The population likelihoods method calculate the ancestral population likelihood from the product of genotype frequencies for each locus. The Plug-in calculates the likelihood for 65 populations based on the genotype frequencies of 151 SNPs. The Population Likelihood analysis is used to calculate the likelihood of the sample genotypes for a particular population according to its genotypes and allele frequencies. The value shows the sub-population to express, which a tested sample is most likely to belong and the likelihood values are listed in order from highest to lowest.

The Admixture prediction and Population likelihood results were viewed based on genotype for all the markers were exported as pdf files as summary reports. The summary includes the Map subtab of the Plug-in and the results are displayed for each sample. The Map subtab was shown as a heat map overlaid on a world map to see for each sample the most likely geographical origin with reference to seven root populations. The heat map colour shows the proportion of the corresponding

population in the sample and a report displays the distribution of log-likelihood values and the confidence level as shown in Figure 5.11 and figure 5.12, respectively.

(A)



(B)

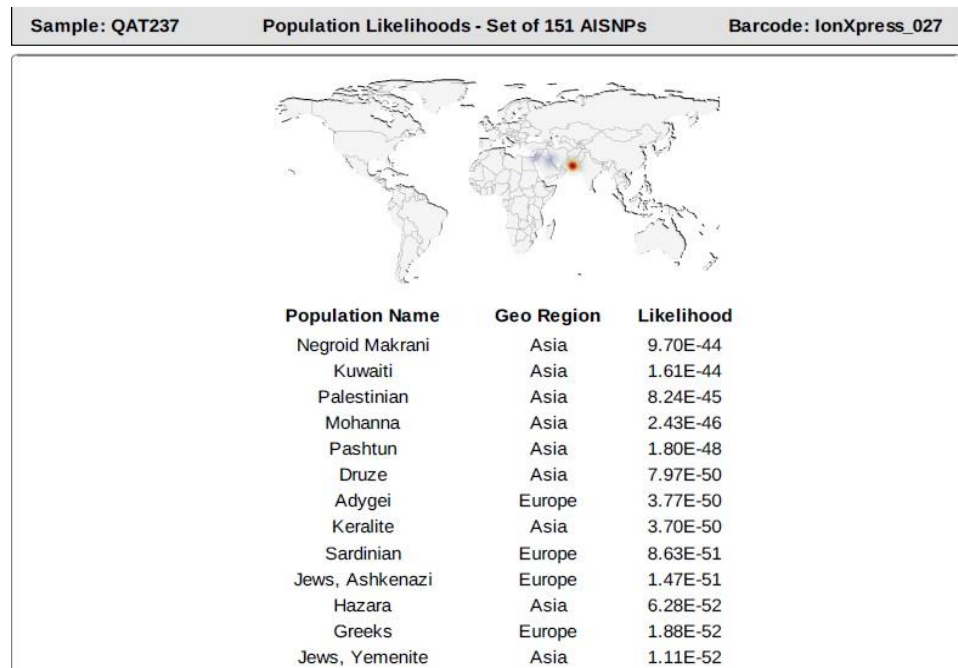
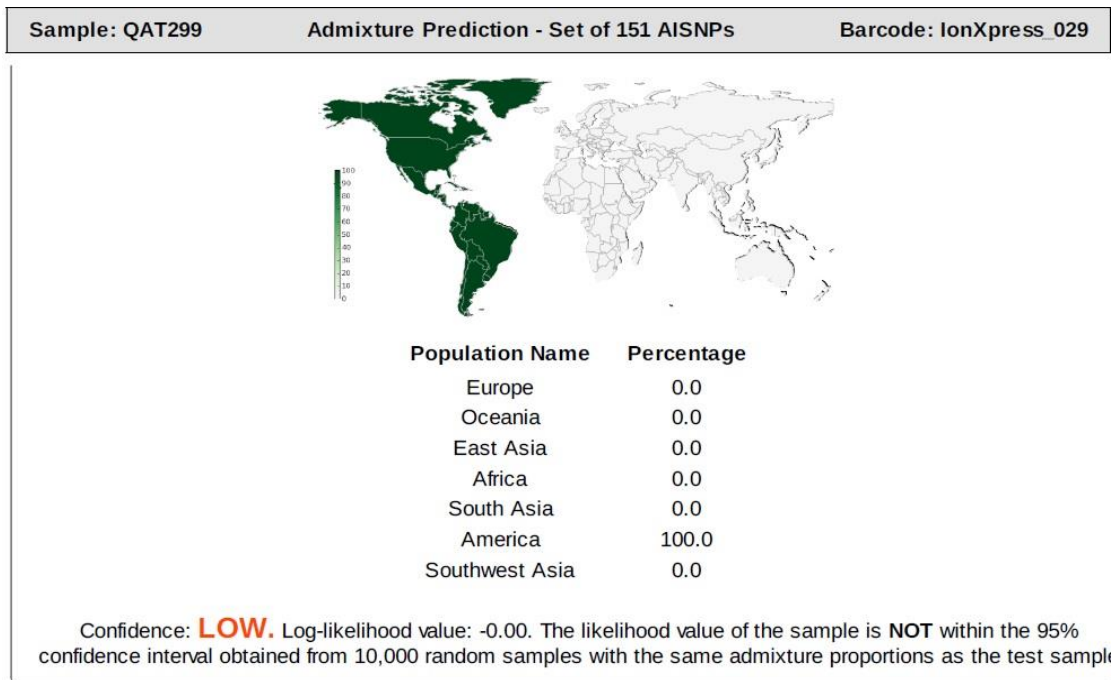


Figure 5.11. Maps and data in (A) and (B) show the Admixture Prediction and the Population Likelihoods for sample QAT237.

(A)



(B)

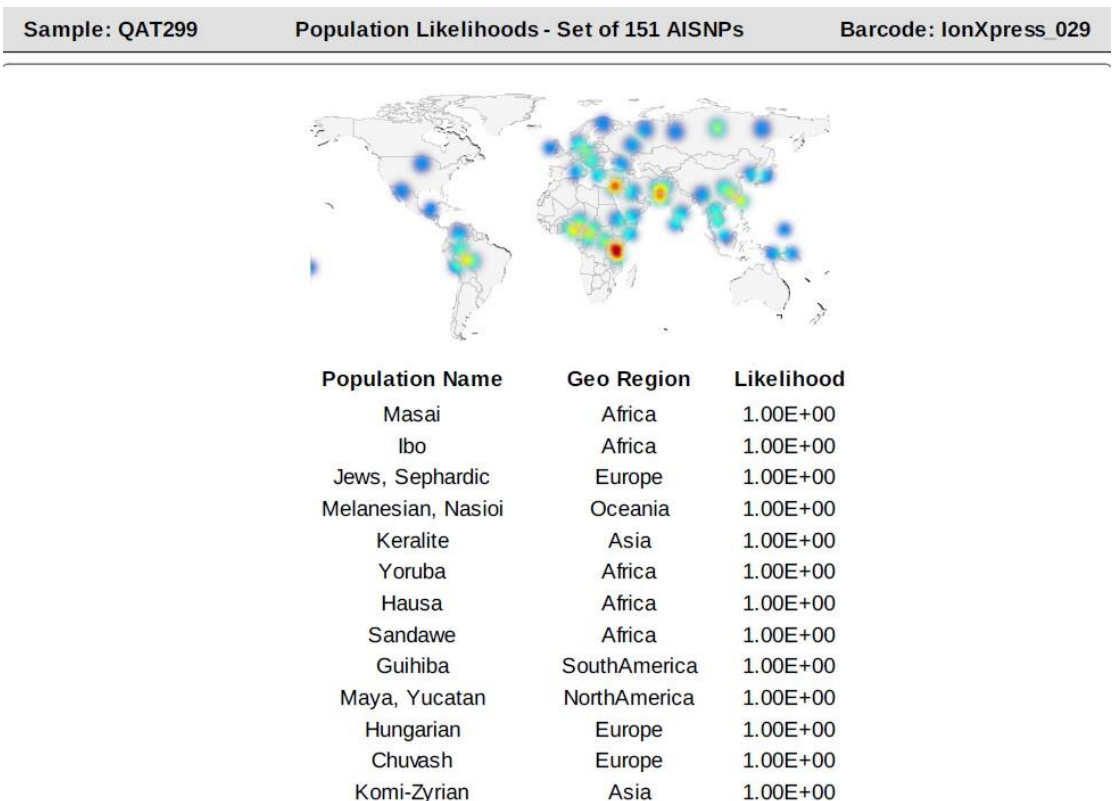


Figure 5.12. Maps and data in (A) and (B) showing sample QAT299 which had no genotype.

An excel file was prepared containing all the sample names, their barcodes, sex, the name of municipality, the result of Admixture Prediction and Population likelihood values (Appendix 14). The Admixture prediction and Population likelihood were obtained for 299 samples, which were collected from the eight municipalities of Qatar (Table 5.8). Unsurprisingly, most of the people from the Qatari population were predicted to have Asian origin, especially Southwest Asia and South Asia ancestry. The results showed that Palestinian are most frequently as the probable population of origin, followed by Kuwaiti and Yemenite Jew. Some samples were reported as most likely to be Druze, Pashtun and Negroid Makrani. Fewer samples were reported to have African ancestry and European ancestry (Table 5.9).

Table 5.8. Data showing the most likely population of origin results predicted of 299 samples from Qatari population were collected from eight municipalities. Each colour corresponds to the color of the municipality presented in Qatar map found the Appendix 1.

Municipality Name	Doha (75)	Al Rayyan (132)	Umm Salal (37)	Al Khor and Al Thakira (12)	Al Daayen (8) and 1 no genotype	Al Wakra (20)	Al Shamal (5)	Al Sheehaniy (10)
Palestinian	26	45	6	3	1	5	2	4
Kuwaiti	13	31	11	5	2	4	2	1
Jews, Yemenite	7	24	5	2	-	2	1	2
Keralite	-	-	1	-	-	-	--	1
Pashtun	12	2	-	-	1	2	-	-
Negroid Makrani	6	5	3	-	1	-	-	-
Druze	5	9	6	2	2	2	-	-
Adygei	1	1	-	-	-	-	-	-
Mohanna	-	-	-	-	-	1	-	-
Greeks	-	-	-	-	-	2	-	-
Jews, Sephardic	3	2	1	-	-	-	-	1
Jews, Ashkenazi	-	1	-	-	-	-	-	-
Sardinian	-	2	2	-	-	-	-	-
Samaritans	-	1	-	-	-	-	-	-
African Americans	1	3	2	-	-	-	-	-
Somali	-	3	-	-	-	1	-	-
Sandawe	-	-	-	-	-	1	-	-
Jews, Ethiopian	1	3	-	-	1	-	-	1

Table 5.9. Data showing the most likely population of origin predicted for the 299 individuals.

Population Names	Total number of individuals
Palestinian	92
Kuwaiti	69
Jews, Yemenite	43
Druze	26
Pashtun	17
Negroid Makrani	15
Jews, Sephardic	7
African Americans	6
Jews, Ethiopian	6
Somali	4
Sardinian	4
Keralite	2
Adygei	2
Greeks	2
Mohanna	1
Jews, Ashkenazi	1
Samaritans	1
Sandawe	1

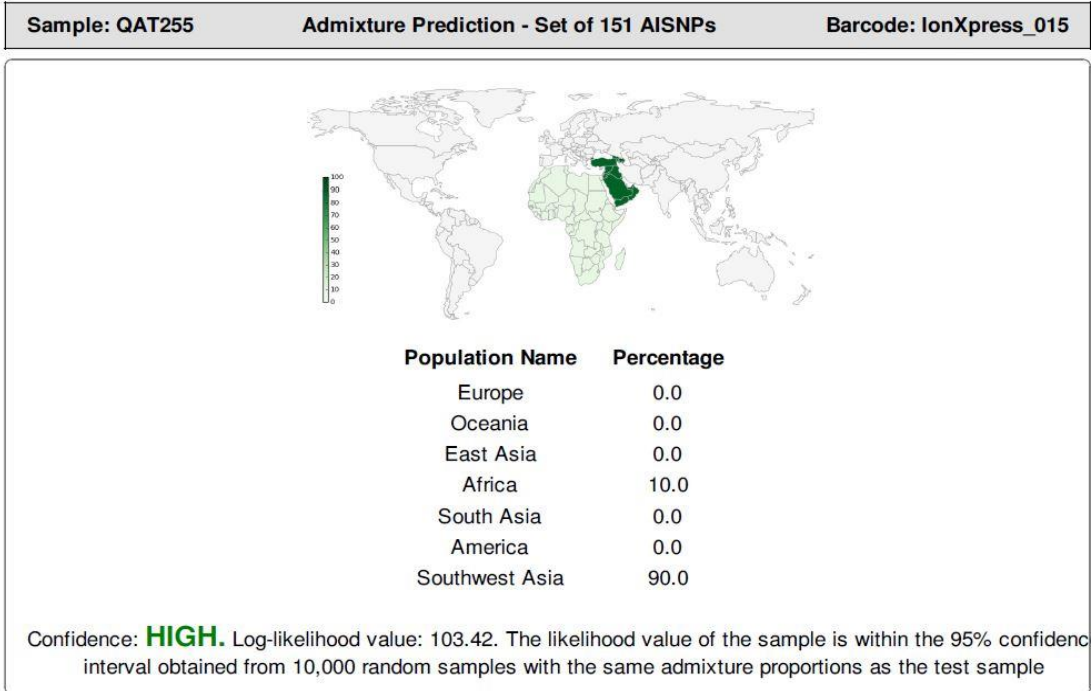
5.5.2 Ancestry prediction with a customized HID SNP Genotyper (Addition of Qatar population data)

One hundred three samples were randomly selected from the 300 Qatari individuals from the eight municipalities that make up the state. By using PowerStat Microsoft Excel Workbook (Tereba 1999), the allele frequencies were calculated to be sent to Thermo Fisher Scientific to design a customized HID SNP Genotyper plug-in. Then, allele frequencies were added to the existing population likelihood as a new group called 'Qatar'. A modified version of the HID_SNP_Genotyper plugin with Qatar data was designed. Running the modified HID SNP Genotyper used Qatari frequencies alongside all the other groups that were already in the plugin. Qatari population samples were then re-analyzed with a modified version of HID_SNP_Genotyper Plug-in; HID_SNP_Genotyper_Qatar_v1. The total number of samples in the data set is 103 Qatari samples. The remaining 197 samples were re-analyzed using the customized plug-in, taking into account the sample that was failed (QAT 299) (it was successfully re-amplified and included within the 197 samples).

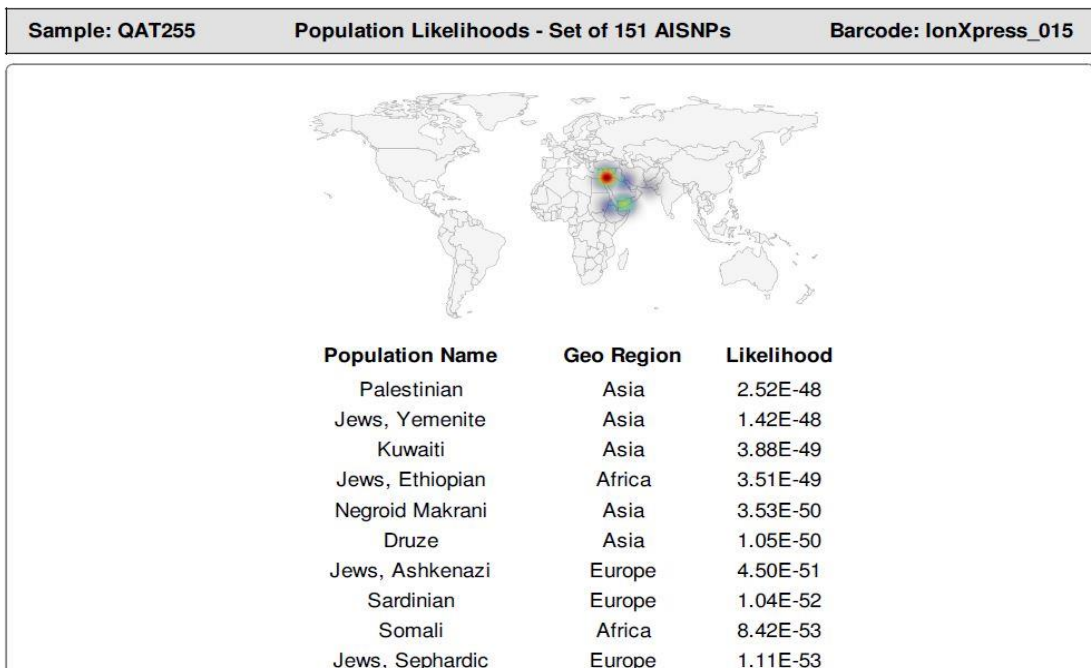
Table 5.10. Data showing the predicted populations for the Qatari individuals studied with Precision ID Ancestry Panel. Column in the middle represents the previous results. The right column represents the results of the modified plug-in with Qatar population likelihood. The green line shows the number of individuals predicted as Qatari. Orange indicates populations with reduced numbers assigned; grey represents no change.

Population Names	Total 299 Qatari samples analyzed with default plugin	*Total 197 Qatari samples re-analyzed with default plug-in	*Total 197 Qatari samples re-analyzed with customized plug-in
Qatar	-	-	91
Palestinian	92	63	19
Kuwaiti	69	43	30
Jews, Yemenite	43	30	11
Druze	26	14	8
Pashtun	17	12	9
Negroid Makrani	15	9	7
Jews, Sephardic	7	5	3
African Americans	6	5	5
Jews, Ethiopian	6	4	4
Somali	4	4	4
Sardinian	4	3	1
Keralite	2	0	0
Adygei	2	1	1
Greeks	2	1	1
Mohanna	1	0	0
Jews, Ashkenazi	1	1	1
Samaritans	1	1	1
Sandawe	1	1	1

(A)



(B)



(C)

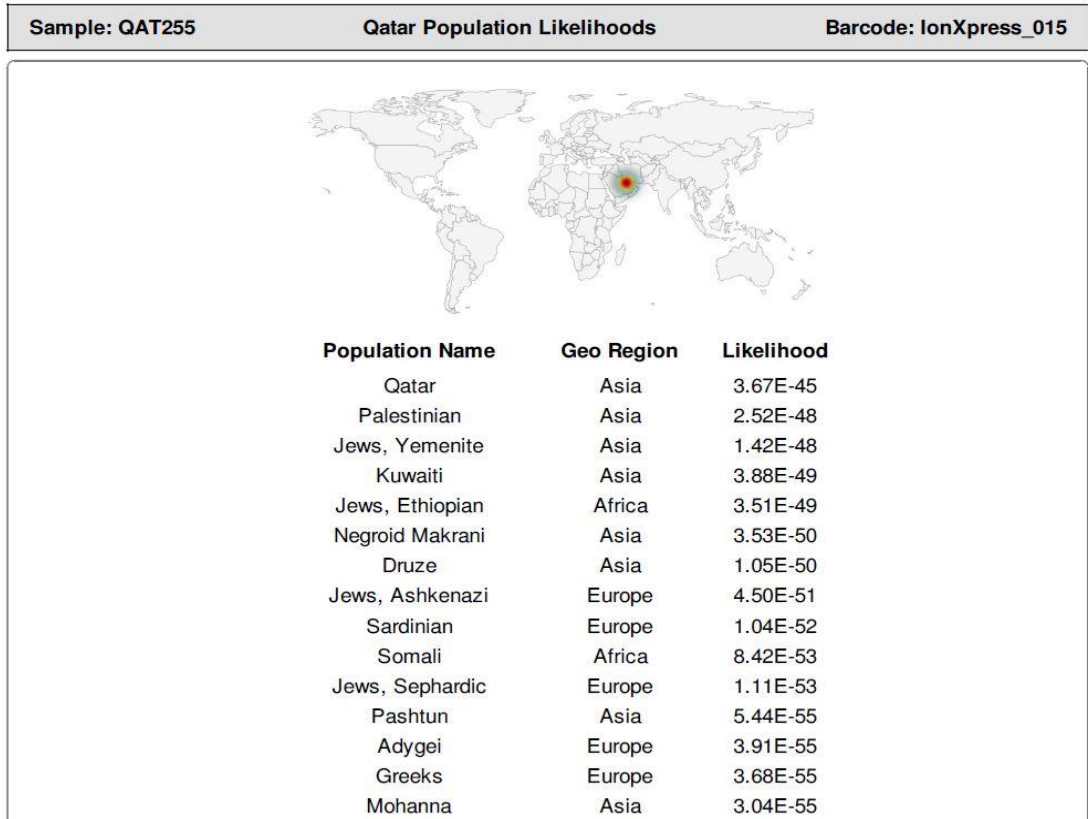


Figure 5.13. Maps and data showing re-analysis results of sample QAT255.(A) Admixture Prediction. (B) Population Likelihoods (default) and (C) Qatar Population Likelihoods (addition Qatar data).

5.5.3 STRUCTURE Analysis

Population structure was assessed through using STRUCTURE 2.3.4 software (Pritchard et al. 2010). The eight municipalities in this study were considered as 8 populations and were analysed using admixture model with correlated allele frequencies and a burn of 10,000 iterations followed by a run of 100,000 iterations. The runs were on k value ranges from k=1 to k=8 with 3 replicates. The results were obtained from STRUCTURE were downloaded onto an online tool STRUCTURE Harvester to calculate the most likely number of K population through the Evanno method.

The optimum K value was generated from STRUCTURE Harvester and the value was K=4, which indicated that four populations were the most likely, based on the 165 SNPs studied. Similarly, results of the structure analysis were visualised and extracted using online web tool "POPHELPER Structure Web App <http://www.pophelper.com/>) (see Figure 5.14 and Figure 5.15). Whilst the most likely number of populations was estimated to be 4, the visualisation of the data illustrates that the populations cannot be clearly identified. As such, the differentiation of populations is very limited and that in contrast to the delta K value the most likely explanation is that it is one population without clear subdivisions.

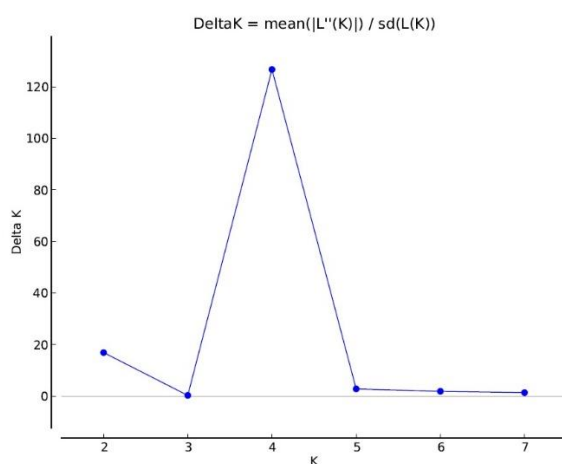
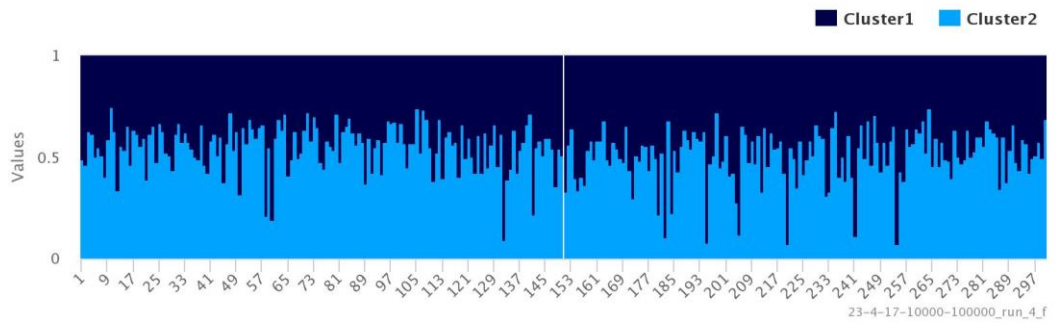
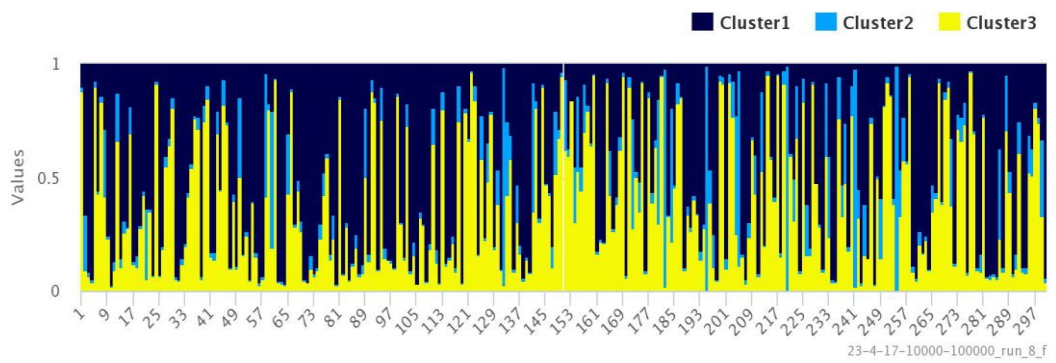


Figure 5.14. The graphical representation of assigning K value based on the Evanno method generated from STRUCTURE-Harvester software for 8 populations.

K=2



K=3



K=4

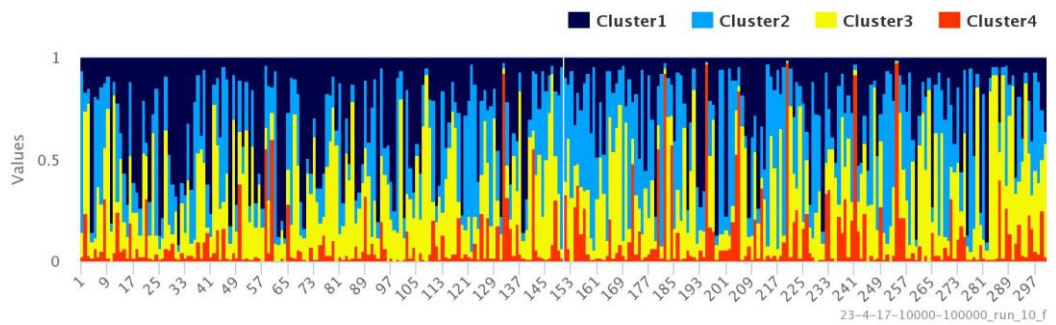


Figure 5.15. Original images of structure analysis results based on 165 SNPs; the most likely number of clusters (K) is K=4.

5.7 Discussion

The ancestry of people currently living the Arabian Peninsula is complex, with multiple waves of migration through the region dating back as far as over 100,000 years. In addition, it is located between three continents namely Africa, Asia Europe, with significant populations movements between Africa and Asia (Omberg et al., 2012). In addition, the population, in common with much of the Arabic world, has a high level of consanguinity. This leads to pockets of communities with relatively high levels of inbreeding (Rodriguez-Flores et al., 2016).

Precision ID Ancestry Panel was applied to test Qatar data population. The software provides the Admixture population and Population Likelihoods results. The Admixture Population detect the racial percentage of a sample from the seven root populations present in the Plug-in. From the sample genotype, the software identifies the best geographical combination from these populations including Africa, America, Middle East (Southwest Asia), Europe, Oceania, East Asia and South Asia. The population Likelihoods detect the population of origin of the sample/an individual based on allele/genotype frequencies from 65 populations in the Plug-in (Pereira et al., 2017). Currently, Qatar population is not included in the database of the default HID SNP Genotyper software.

In this study, the collection of 300 samples from the eight municipalities was used to assess geographical distribution and the diversity of Qatar population in Qatar using the HID Ancestry Panel, whereas mentioned, the people came to Qatar across the Qatari mainland and across the Eastern coast of the Arabian Gulf (Obaidan, 1982). The region of the modern country of Qatar has been at the crossroads of major migrations from the eras of ancient humans, early civilization and recent centuries (Omberg et al., 2012).

The results obtained from this study which were most of Qatar population were predicted to have Southwest Asian and South Asian ancestry, with some samples predicted to have African ancestry and European ancestry, although the ancestral populations identified in Africa and Europe are of Middle Eastern origin, e.g. Jewish ancestry. The results explained and supported the historical migration, settlement and trade theory in the country.

Progress was seen in this study with addition of a set of data indicating the allele frequencies of Qatari samples to design a customized HID SNP Genotyper plug-in. Once the reference data of Qatari population were included, the samples were re-analyzed with a modified version of the Plug-in "HID_SNP_Genotyper_Qatar_v1". The following changes were seen and they included the 86.8% of the samples that were predicted to be Asian in origin. They are believed to be originated from Southwest Asia and with South Asia Ancestry representing 88.83%. The samples which were predicted to have Europe ancestry comprised only 4.06%, with 2.03% of the samples reassigned to have Southwest Asian ancestry instead of Europe ancestry (Figure 5.16); but as noted earlier the African and European populations are of Middle Eastern origins in many cases.

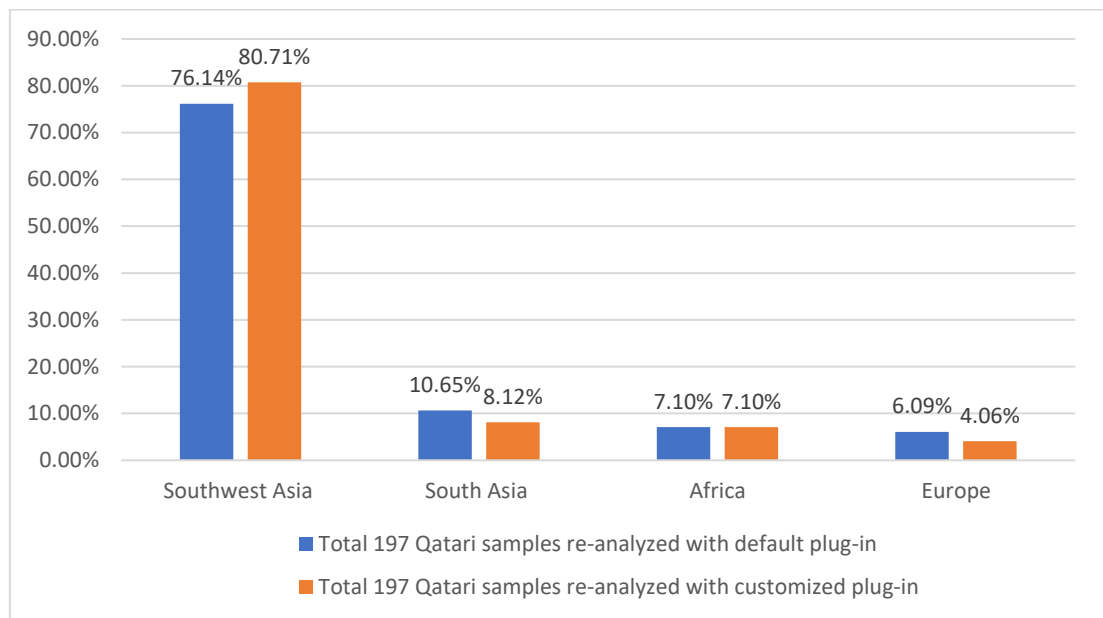


Figure 5.16. Bar charts showing Geographic distribution for 197 Qatari samples analyzed with default and customized HID SNP Genotyper plugin.

The samples with estimated European ancestry in the study are from the populations neighboring the Arabian Peninsula and the Middle East and close to the ancient migration paths. The Arabian Peninsula was the original homeland of Semites; following the advent of Islam and also through conflict populations spread to neighboring regions and countries (Shamsuddin & Ahmad, 2020).

STRUCTURE software was used to study the distribution of genetic ancestry in each sample. The results indicate an optimum K=4 was observed based on the 165 SNPs studied, which indicated that four populations were the most likely; joining to group

the populations into 4 clusters. The plot in the Figure 5.14 and Figure 5.15 of this chapter and mentioned above displays each individual as a thin vertical bar with colour representing the membership proportion of individual to cluster. STRUCTURE tries to fit each individual into the best cluster. However, the figure whether $k=3$ or $k=4$ showed that the majority of Qatari individuals are intermediate between all the clusters. Moreover, the cluster cannot allocate each individual into a single group. Population structure of Qataris in comparison to neighbouring and/ or world populations might be more revealing.

The main genetic features of the Qatari population seem to derive from at least two major factors. One aspect is the high rate of marriages between close biological relatives that characterizes this Arab population. The other is its strategic geographic position in the Arabian Peninsula close to the Horn of Africa migratory passageway, an important pivotal contact zone for bidirectional migrations between Eurasia and Africa (Pérez-Miranda et al., 2006).

In summary to this section, the addition of Qatari reference data to this study was aimed to improve the results obtained from this dataset and any further investigations that will be held in the future and depend on this panel in the Qatar population. The improvement was seen in some of the results that were assigned to Southwest Asia. It does also illustrate the limitations of the reference datasets that are available.

5.8 Conclusion

For the first time, the Precision ID Ancestry Panel was applied to the population of the State of Qatar by studying samples consisting of 300 Qataris from the eight municipalities that make up the State. The sequencing performance, in term of coverage, forensic parameters and ancestry inference ability of the panel, was investigated and the sensitivity experiment. The panel is performed well and high informative assay for ancestry population prediction. In general, this new MPS technology and the results obtained can be a useful tool for forensic human identification and biogeographical inference.

The panel can be applied as a supplementary instrument with the routine STR markers especially in complex paternity cases or with challenged casework samples. However, given the genetic similarities between Qatar and its neighboring countries, there is a

limited scope for the application of the technology. As demonstrated, the addition of the Qatari data and the availability of relevant population databases are important in trying to identify ancestral populations. More work could be undertaken to develop a higher resolution overview of the Qatari population, but more important is the generation of reference populations for the regional region. The overall effectiveness of the ancestry markers will be discussed in more detail in Chapters 6 and 7.

CHAPTER: 6 Analysis of forensic casework samples by Precision ID Ancestry Panel – Manual and Automated AmpliSeq Workflow

6.1 Introduction

Studying ancestry can potentially be useful in many fields including forensic casework analysis. Inference ancestry in forensic analysis is important by adding additional and useful information to the case file. Such cases involved the characterization by challenging cold cases or unsolved ones when there is no match found between the sample and any living person or available database records. For example, in the 2004 Madrid train bombings, source population of the suspects was inferred by using 34 autosomal SNPs related to the ancestry of population (Phillips et al., 2009; Ziętkiewicz et al., 2012).

The data presented in Chapter 5 illustrates the potential of the Precision Ancestry panel to identify the geographical origins of samples self-identified as Qatari, and also highlighted the limitations in the availability of reference databases.

This chapter was designed to evaluate the performance and suitability of the Precision ID Ancestry Panel on actual casework samples selected from real cases.

Objectives

1. To collect different casework samples from original cases from different security departments which represent eight municipalities that make up State of Qatar.
2. To evaluate the following: -
 - a) NGS performance; massively parallel sequencing workflow as a tool in forensic casework analysis.
 - b) Automated library preparation performance.
 - c) Precision ID Ancestry panel in different casework samples that were extracted with different extraction protocols used in Qatar Forensic Laboratory.
 - d) The Panel's (165 SNPs) ability to predict ancestry in unknown crime scene samples with different DNA concentration.

6.2 Materials and Methods

A total of 143 casework samples were selected for this study from 76 real cases, which had both partial and full genotypes. The samples collected between 2005 and 2018 and were profiled previously with Identifiler[®], Identifiler[®] Plus, MiniFiler[™], GlobalFiler[®]. Some samples were amplified with Yfiler[®] and Yfiler Plus[®] (Applied Biosystems[™]).

The materials and methods relevant to this Chapter were described in Chapter 2.

Part one: Manual Library Preparation Workflow

6.3 Results

One hundred and eleven casework samples were selected from different biological material including hair, cells, semen, saliva, blood, touch, teeth and bone. The samples were available as extracted DNA. Library construction was performed using the Precision ID Ancestry Panel and the Ion Precision ID Library Kit (TFS). The samples with DNA concentration of 1ng and above were amplified with 21 cycles whereas for samples with less than 1ng available, 26 cycles of amplification were used. An Ion Chef™ was used for template preparation and chip loading. Templated amplicons were sequenced using Ion Torrent PGM instrument (TFS) on 316 v2 Chips on 5 chips (Table 6.1).

Table 6.1. Data showing the total number of casework samples were used in manual library preparation workflow.

Sample Types	Chip1	Chip2	Chip3	Chip4	Chip5
Hair	1	0	1	1	0
Cells	8	11	5	5	1
Semen	1	1	1	0	1
Bone	0	3	0	2	1
Teeth	0	1	0	1	0
Blood	2	0	9	0	12
Saliva	1	0	6	3	3
Touch	1	8	2	7	1
Tissue	10	0	0	1	0
Total	24	24	24	20	19

6.3.1 Quality of the sequencing runs

The quality for each run was evaluated through the output reports generated from Torrent Suite™ Software 5.4.0, through Summary & Quality control metrics.

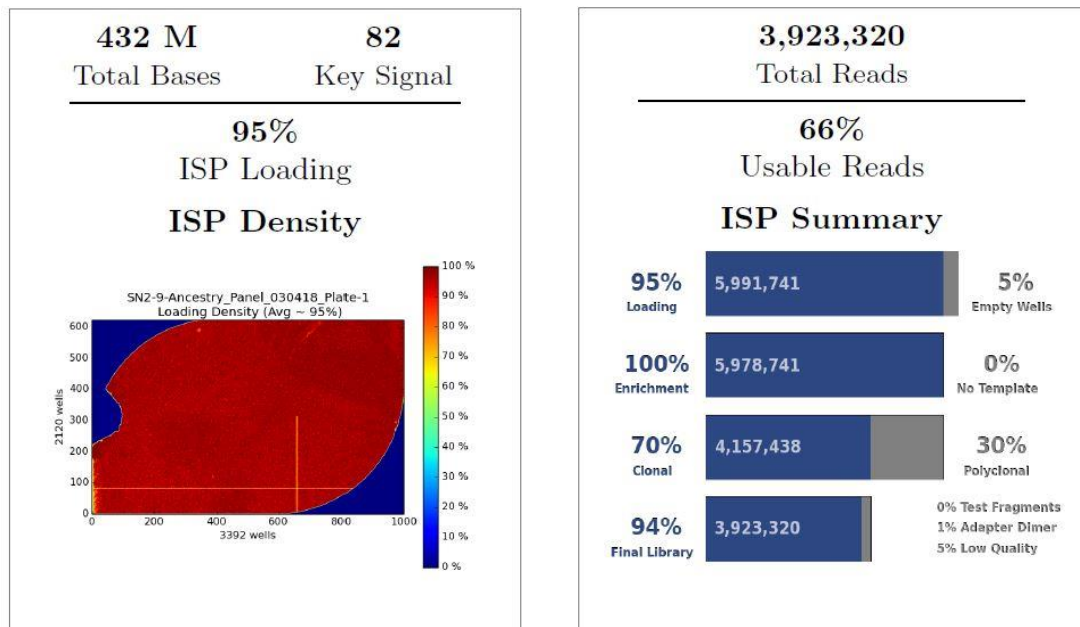
The results showed that the highest ISP loading percentage was seen in Chip#1 followed by Chip#5 and Chip#3 with 95%, 93% and 91%, respectively. Whereas Chip#4 and Chip#2 had 89% and 81% each of which contained the weaker samples that had been amplified for 26 cycles (Table 6.2).

Table 6.2. Data showing the average ISP density summarized for each chip used.

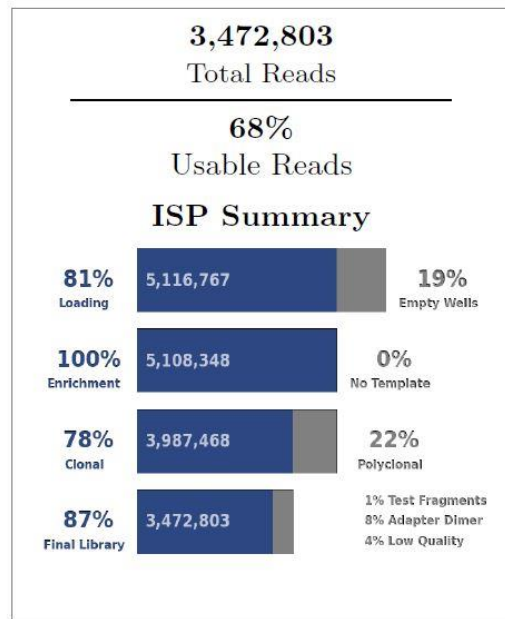
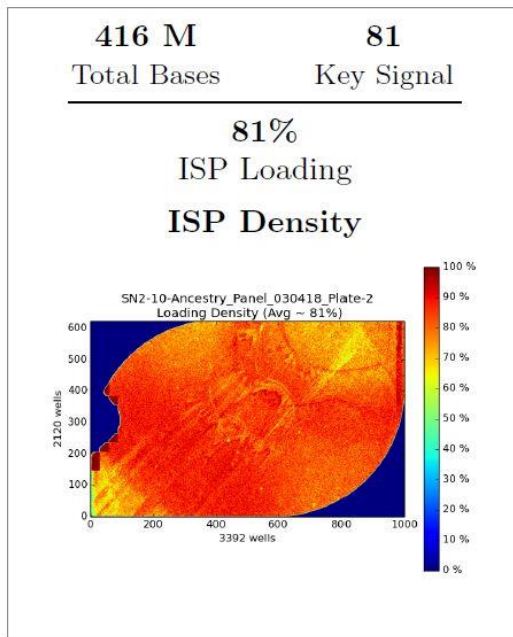
Chip#	Number of samples	ISP loading %
1	24	95
2	24	81
3	24	91
4	24	89
5	19	93

Also, the quality of the ISPs that were loaded onto the sequencing chip was shown in the second part which was ISP summary. The mean read length for all the 5 chips were: 110 bp, 120 bp, 103 bp, 131 bp, 118 bp, respectively (see Figure 6.1).

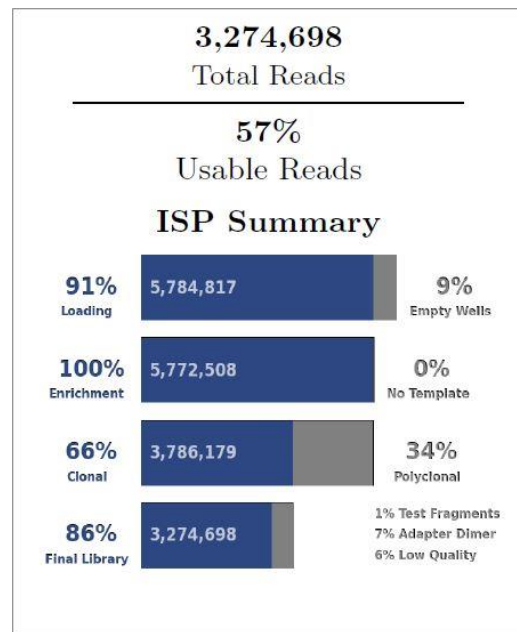
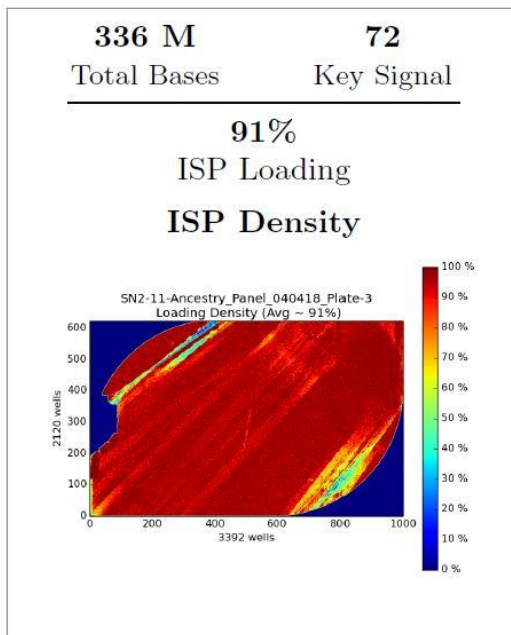
Chip#1



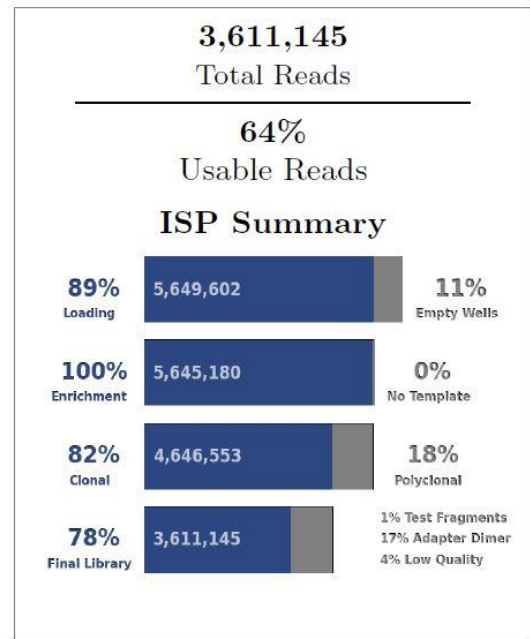
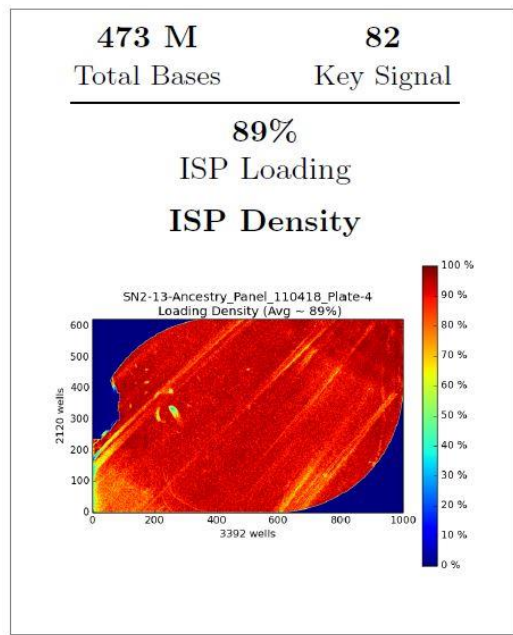
Chip#2



Chip#3



Chip#4



Chip#5

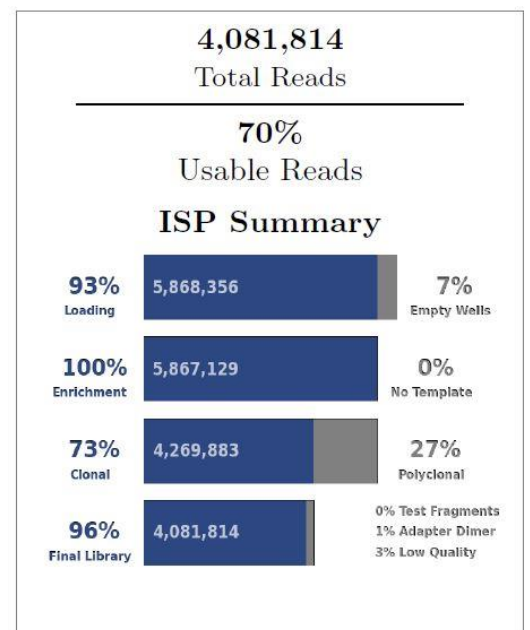
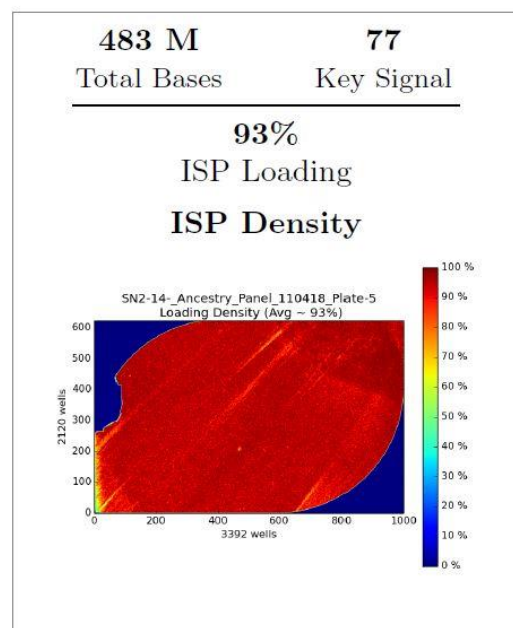


Figure 6.1. Diagrams showing Chips 1-5 displaying the ISP density percentage and ISP summary for Chip#1, Chip#2, Chip#3, Chip#4 and Chip#5 which were taken from summary report for each chip.

6.3.2 Sequence Coverage

The mean depth of coverage was obtained from the coverage analysis report as shown in Table 6.3, for all Ion 316 Chips that were used to sequence casework samples.

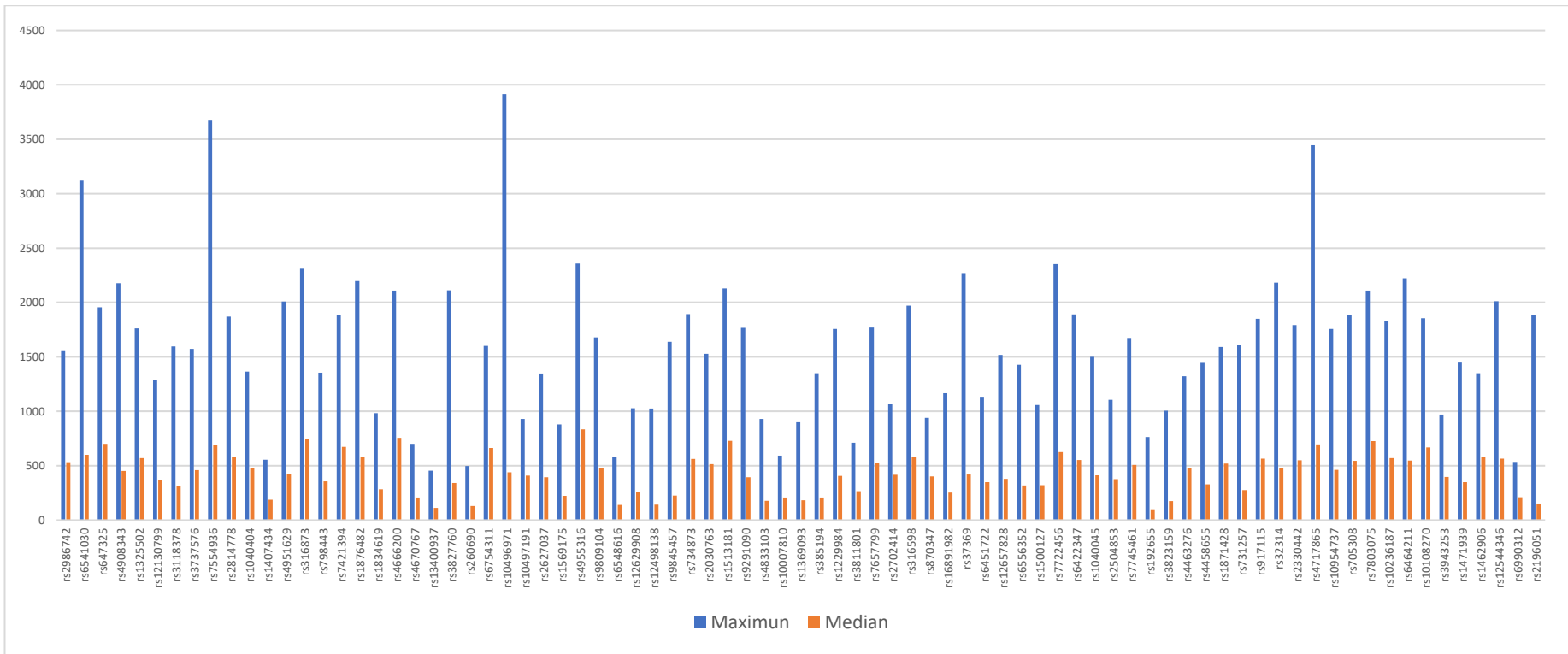
Table 6.3. Table showing the mean depth for the 5 Chips used in this study.

Chips#	Averages of Mean Depth
1	559.97
2	241.70
3	466.43
4	95.31
5	837.08

The average coverage across the 165 SNPs was 447x. The maximum coverage value was observed with SNP, rs3907047 (4440x). Figure 6.2 shows the maximum and median values for each SNP from all the casework samples, the average median value was 464.95x with range from 50x (rs7997709) to 879x (rs1760921). Also, Table 6.4 shows the lowest coverage values that were observed and shown among all samples.

Table 6.4. Data showing the SNPs with the lowest coverage.

SNP IDs	number of samples	Chip#
rs1296819	20	1,2,4,5
rs260690	13	1,5
rs192655	9	3
rs2986742	8	3,4
rs2196051	7	2,3,4
rs7997709	6	2,3
rs1950993	5	5
rs12498138	5	1,2,3,4
rs1407434	4	1,3,5
rs12130799	3	1,2,4
rs310644	3	3,4
rs1325502	2	2,4
rs9845457	2	3
rs4670767	2	4,5
rs10007810	1	2
rs4463276	1	1
rs200354	1	2
rs4908343	1	2
rs2504853	1	2
rs4458655	1	2
rs3118378	1	2
rs647325	1	2
rs13400937	1	5
rs6541030	1	4
rs1569175	1	4



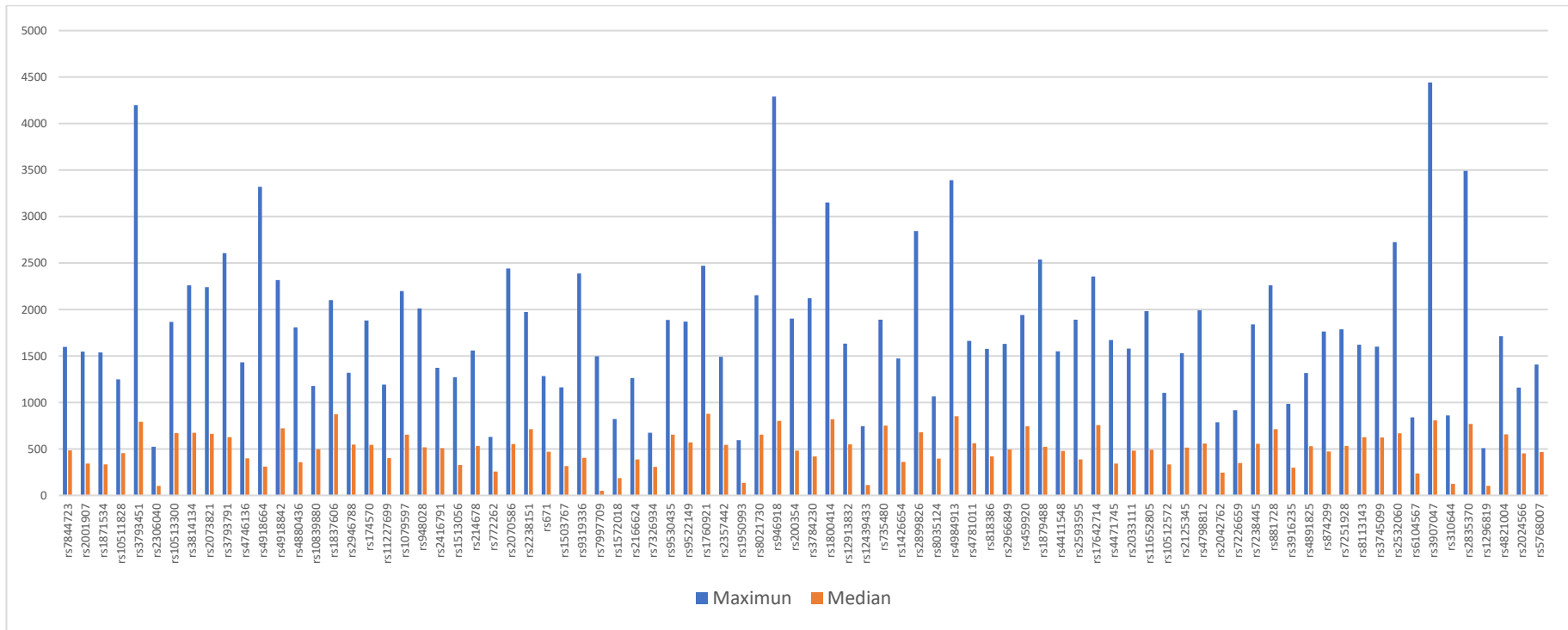


Figure 6.2. Column charts showing maximum and median of the coverage for each SNP included in the Precision ID Ancestry panel resulted from casework samples (manual library preparation experiment).

6.3.3 Performance of the Precision ID Ancestry Panel

6.3.3.1 Full SNP profile

Full SNP profiles were successfully generated from 47 samples and they include multiple biological evidence and some of which were characterized as challenged samples (Table 6.5). The yellow row in the table represents the full successful profile generated from the lowest DNA concentration (i.e. 0.02 ng).

Table 6.5. Data showing the full SNP profiles were seen in 47 casework samples. Columns shaded brown contained suboptimal DNA template.

Samples#-Barcode#-Plates#	Year of Analysis	Quantification values (ng/ μ l)*	Sample Types	STR results
S4-BC4-P1	2005	10.34	Cells collected from teeth cleaning twig "Miswak"	PPlus, Re-amp-GF/FP
S5-BC5-P1	2005	19.46	Cells collected from tissue paper)	PPlus, Re-amp-GF/FP
S6-BC6-P1	2005	273.98	Tissue	PPlus, Re-amp-GF/FP
S7-BC7-P1	2005	679.32	Tissue	PPlus, Re-amp-GF/FP
S9-BC9-P1	2005	162.09	Tissue	PPlus, Re-amp-GF/FP
S11-BC11-P1	2005	122.44	Tissue	PPlus, Re-amp-GF/FP
S12-BC12-P1	2005	162.76	Tissue	PPlus, Re-amp-GF/FP
S13-BC13-P1	2005	258.91	Tissue	PPlus, Re-amp-GF/FP
S14-BC14-P1	2005	522.82	Tissue	PPlus, Re-amp-GF/FP
S15-BC15-P1	2005	271.54	Tissue	PPlus, Re-amp-GF/FP
S16-BC16-P1	2017	12.25	Touch (swab from bullets)	GF/FP
S18-BC18-P1	2017	0.24	Hair root	IDP/FP
S19-BC19-P1	2017	12.18	Saliva from cigarettes	IDP/FP
S20-BC20-P1	2017	5.10	cells from tissue paper	IDP/FP
S22-BC22-P1	2017	0.57	Blood stain from the ground	GF/FP
S24-BC24-P1	2018	25.79	Blood stain blade knife	GF/FP

Samples#-Barcode#-Plates#	Year of Analysis	Quantification values (ng/μl)*	Sample Types	STR results
S5-BC5-P2	2017	0.1 [^]	Cells. (Swab from Deodorant)	IDP/FP
S10-BC10-P2	2005	0.02 [^]	Cells. (swab from a cup)	PPlus/PP
S17-BC17-P2	2018	0.06 [^]	Cells. (Swab purple plastic cup)	GF/FP
S1-BC1-P3	2017	1.3	Cells from tooth stick	IDP/FP
S2-BC2-P3	2017	0.98	Cells from bandage	IDP/FP
S3-BC3-P3	2017	118.10	Blood stain	GF/FP
S4-BC4-P3	2017	116.48	Swab-Blood stain	GF/FP
S8-BC8-P3	2016	0.43	Saliva-Reference	NT
S9-BC9-P3	2016	27.90	Saliva stain	IDP/FP
S10-BC10-P3	2017	37.10	Blood stain	GF/FP
S11-BC11-P3	2017	3.57	Blood stain	GF/FP
S12-BC12-P3	2016	11.52	Semen stain (Sperm fraction)	IDP, Re-amp-GF/FP
S13-BC13-P3	2016	11.75	Blood stain	IDP, Re-amp-GF/FP
S20-BC20-P3	2017	4.28	Blood stain	GF/FP
S24-BC24-P3	2017	17.21	Saliva stain	IDP, Re-amp-GF/FP
S2-BC2-P5	2016	3.29	Blood stain	IDP, Re-amp-GF/FP
S3-BC3-P5	2016	14.20	Blood stain	IDP, Re-amp-GF/FP
S4-BC4-P5	2016	3.53	Blood stain	IDP, Re-amp-GF/FP
S5-BC5-P5	2016	18.89	Blood stain	IDP, Re-amp-GF/FP
S6-BC6-P5	2016	52.39	Blood stain	IDP, Re-amp-GF/FP
S7-BC7-P5	2016	16.28	Blood stain	IDP, Re-amp-GF/FP
S8-BC8-P5	2018	4.09	Cells from cutting towel	GF/FP
S9-BC9-P5	2017	11.00	Saliva stain	GF/FP
S10-BC10-P5	2017	0.94	Blood stain	GF/FP

Samples#-Barcode#-Plates#	Year of Analysis	Quantification values (ng/μl)*	Sample Types	STR results
S13-BC13-P5	2015	17.02	Bone	IDP, Re-amp-GF/FP
S14-BC14-P5	2017	27.42	Blood stain	GF/FP
S15-BC15-P5	2017	1.21	Blood stain	GF/FP
S16-BC16-P5	2018	3.49	Blood stain	GF/FP
S17-BC17-P5	2017	26.52	Semen stain (Sperm fraction)	GF/FP
S18-BC18-P5	2018	2.9	Saliva from cigarette	GF/FP
S19-BC19-P5	2018	1.9	Blood stain	GF/FP

* All DNA sample concentration was adjusted to 1 ng and volumes based on quantity of DNA. ^ The maximum volume of extracts (6 μl) was added in case of some samples yielded concentrations of DNA below that value. FP- Full profile, PP- Partial profile, NT- Not tested, GF- Global Filer, IDP- Identifiler Plus and PPlus- Profiler plus. Re-amp-sample reamplified; according to the extract volume available in the sample tube, some samples were reamplified with GlobalFiler.

6.3.3.2 Partial SNP profiles

As shown below in Table 6.6, incomplete SNP genotype profiles were observed from 53 casework samples at different proportions across all 165 SNPs. Eight samples generated profiles from 164 SNPs, wherein, one SNP was not genotyped from these samples (different SNPs dropped out). This is with an average of 99.39% of profile recovery (cells highlighted in green). Out these eight samples, 1 sample had DNA concentration of less than 1 ng and the remaining 7 samples were above 1 ng. Partial profiles with 2-5 SNP missing genotype data were in 8 samples (cells highlighted in blue). Their average profile recovery was 98.18%. Nine samples generated partial profiles and the number of SNP genotype missed in those samples were 6-10 (highlighted in yellow).

Five samples showed partial genotypes with between 12-15 drop-outs; the profile recovery percentage was 92.73%, 92.12%, 91.52% and 90.91%, respectively (highlighted in pink). The partial profiles collected from remaining samples had

different rates of drop-out, with the highest dropout with 164 SNPs not profiled in a bone sample received in 2016; the DNA quantification value was 0.01 ng (highlighted in grey).

Table 6.6. Results showing the partial SNP profiles were seen in 53 casework samples.

Samples#- Barcodes#- Plate#	Year of Analysis	DNA Concentrations (ng)*	Sample Types	#Drop out SNPs	Profile recovery (%)	STR results
S3-BC3-P1	2005	0.85	Cells collected from razor Blade	1	99.39	PPlus/FP
S5-BC5-P3	2017	14.34	Blood stain	1	99.39	GF/FP
S15-BC15- P3	2017	6.65	Blood stain	1	99.39	IDP/FP
S21-BC21- P3	2016	3.31	Saliva from cigarette	1	99.39	IDP/FP
S22-BC22- P3	2016	2.5	Saliva from cigarette	1	99.39	IDP/FP
S1-BC1-P5	2016	10.21	Saliva from blanket	1	99.39	IDP/FP
S11-BC11- P5	2017	3.34	Touch. Swab from knife handle	1	99.39	GF/FP
S12-BC12- P5	2017	20.42	Blood stain	1	99.39	GF/FP
S1-BC1-P1	2005	1.03	Cells collected from Shirt	3	98.18	PPlus/FP
S8-BC8-P1	2005	76.44	Tissue	2	98.79	PPlus/FP
S21-BC21- P1	2017	1.98	Cells. Swab from pillowcase	5	96.97	GF/FP
S23-BC23- P1	2017	14.64	Semen Stain from condom, sperm fraction	5	96.97	GF/FP
S8-BC8-P2	2005	0.01^	Cells. Swab from chair	5	96.97	PPlus/PP
S9-BC9-P2	2005	0.01^	Cells. Swab from a cup	5	96.97	PPlus/PP
S24-BC24- P2	2018	0.02^	Touch. Swab from a saw	3	98.18	GF/FP
S2-BC2-P4	2015	0.01	Tissue	4	97.58	IDP, Re- amp- GF/FP
S2-BC2-P2	2017	0.01^	Touch. Swab from gun trigger	8	95.15	GF/PP
S3-BC3-P2	2017	0.01^	Teeth	126	23.64	GF/PP
S4-BC4-P2	2017	0.05^	Semen stain. Sperm fraction	10	93.94	IDP/FP
S11-BC11- P2	2005	0.05^	Cells. Swab from a cup	6	96.36	PPlus/PP

Samples#- Barcodes#- Plate#	Year of Analysis	DNA Concentrations (ng)*	Sample Types	#Drop out SNPs	Profile recovery (%)	STR results
S15-BC15- P2	2018	0.06^	Touch. Swab from Steering wheel	7	95.76	GF/FP
S18-BC18- P2	2016	0.01^	Touch swab from Knife handle	10	93.94	IDP/FP
S21-BC21- P2	2017	0.01^	Cells collected from a glove	9	94.55	IDP/PP
S18-BC18- P3	2018	0.61	Cells collected from traditional male headdress.	9	94.55	GF/FP
S1-BC1-P4	2016	0.13^	Touch swab from scissors handle	10	93.94	IDP/FP
S16-BC16- P4	2018	0.32	Cells from tissue paper	7	95.76	GF/FP
S1-BC1-P2	2017	0.01^	Touch. Swab from gun handle	15	90.91	GF/FP
S7-BC7-P2	2005	0.02^	Cells collected from toothbrush	13	92.12	PPlus/FP
S14-BC14- P2	2018	0.06^	Touch. Swab from Steering wheel	14	91.52	GF/FP
S22-BC22- P2	2017	0.01^	Cells collected from a glove	12	92.73	IDP/PP
S6-BC6-P4	2018	0.06^	Touch-from window	12	92.73	GF/FP
S10-BC10- P1	2005	16.77	Tissue	36	78.18	PPlus/FP
S17-BC17- P1	2017	0.20^	Cells. Swab from rope	91	44.85	IDP/FP
S12-BC12- P2	2016	0.01^	Bone	30	81.82	GF/FP
S13-BC13- P2	2016	0.01^	Bone	159	3.64	GF/PP
S19-BC19- P2	2017	0.07^	Touch. Swab from Steering wheel	54	62.27	GF/PP
S20-BC20- P2	2017	0.05^	Cells. Swab from a bottle	139	15.76	GF/FP
S23-BC23- P2	2018	0.01^	Bone	83	49.70	GF/PP
S7-BC7-P3	2017	0.53	Touch. Swab from steering wheel	98	40.61	GF/FP
S16-BC16- P3	2017	0.17^	Touch. Swab from car door back right side	163	1.21	GF/FP
S23-BC23- P3	2017	0.51	Blood stain	127	23.03	IDP/FP

Samples#- Barcodes#- Plate#	Year of Analysis	DNA Concentrations (ng)*	Sample Types	#Drop out SNPs	Profile recovery (%)	STR results
S3-BC3-P4	2018	0.03 [^]	Touch-swab from watch	149	9.70	GF/FP
S4-BC4-P4	2018	0.01 [^]	hair root	37	77.58	GF/FP
S5-BC5-P4	2018	0.07 [^]	Cells. Swab from a cup	65	60.61	GF/FP
S7-BC-P4	2018	0.05 [^]	Touch-safe box	142	13.94	GF/FP
S8-BC8-P4	2018	0.06 [^]	Touch-safe box handle	37	77.58	GF/FP
S10-BC10- P4	2016	0.02 [^]	Cells collected from a glove	99	40.00	IDP/FP
S11-BC11- P4	2016	0.12 [^]	Saliva. Chewing gum	36	78.18	IDP, Re- amp- GF/FP
S12-BC12- P4	2018	0.5	Cells collected from a glove	161	2.42	GF/FP
S15-BC15- P4	2016	0.34	Saliva from cigarette	159	3.64	IDP/FP
S17-BC17- P4	2018	0.2	Touch. Swab from screw handle	55	66.67	GF/FP
S18-BC18- P4	2017	0.14 [^]	Touch. Swab from a gun	128	22.42	GF/FP
S19-BC19- P4	2016	0.01 [^]	Bone	164	0.61	GF/FP

* All DNA sample concentration was adjusted to 1 ng and volumes based on quantity of DNA. [^] The maximum volume of extracts (6 µl) was added in case of some samples yielded concentrations of DNA below that value. FP- Full profile, PP- Partial profile, GF- Global Filer, IDP- Identifiler Plus and PPlus- Profiler plus. Re-amp- sample reamplified; if sufficient volume available, some samples were reamplified with GlobalFiler.

6.3.3.3 No SNP profiles

Out of 111 casework samples, 11 samples failed to generate profiles (Table 6.7).

Table 6.7. Data showing the 11 samples that failed to generate any profile.

Sample#- Barcode#- Plate#	Year of Analysis	DNA Concentration (ng)*	Sample Type	STR result
S2-BC2-P1	2005	0.23	Cells (Swab from Sandals)	PPlus/FP
S6-BC6-P2	2005	0.13 [^]	Cells (Swab from Sandals)	PPlus/PP
S16-BC16-P2	2017	0.04 [^]	Touch (Swab from Knife handle)	GF/PP
S6-BC6-P3	2016	0.29	Cells (Cells scratch from wool hat)	IDP, Re-amp-GF/FP
S14-BC14-P3	2015	0.69	Saliva (Saliva chewing gum)	FP
S17-BC17-P3	2018	0.25	Hair root.	GF/FP
S19-BC20-P3	2018	0.62	Cells (Swab from red plastic cup)	GF/PP
S9-BC9-P4	2016	0.16 [^]	Cells (from a glove)	IDP/FP
S13-BC13-P4	2016	0.01 [^]	Bone	GF/PP
S14-BC14-P4	2018	0.01 [^]	Teeth	GF/PP
S20-BC20-P4	2016	0.7	Saliva-chewing gum	IDP, Re-amp-GF/FP

* All DNA sample concentration was adjusted to 1 ng and volumes based on quantity of DNA. [^] The maximum volume of extracts (6 µl) was added in case of some samples yielded concentrations of DNA below that value. FP- Full profile, PP- Partial profile.

Part two: Automated Library Preparation Workflow

6.4 Results

A total of 47 casework samples and 1 sample from the normal population that was analysed in the previous chapter and failed to generate a profile, were selected to study the automated workflow of library preparation (Table 6.8).

Table 6.8. Data showing the different types of samples were used in automation library workflow.

Sample types	Number of samples
Hair	1
Cells	7
Semen	2
Bone	3
Teeth	1
Blood	9
Saliva	15
Touch	9
Tissue	1

The Ion Chef™ system was used for the automated library preparation using Precision DL8 kit with Precision ID Ancestry Panel (TFS) employing 22 cycles with DNA concentration ≥ 1 ng and 25 cycles for samples <1 ng. For the amplification samples, the concentration was adjusted to 1 ng and the volumes based on the quantity of DNA. In the automation workflow the sample volumes can be increased to 15 μ l according to manufacturer's guidelines (Table 6.9). A total of eight libraries were constructed per run. Template preparation and chip loading was performed on the Ion Chef™ system and then sequencing of loaded 316 v2 Chips was conducted using Ion Torrent PGM system (TFS).

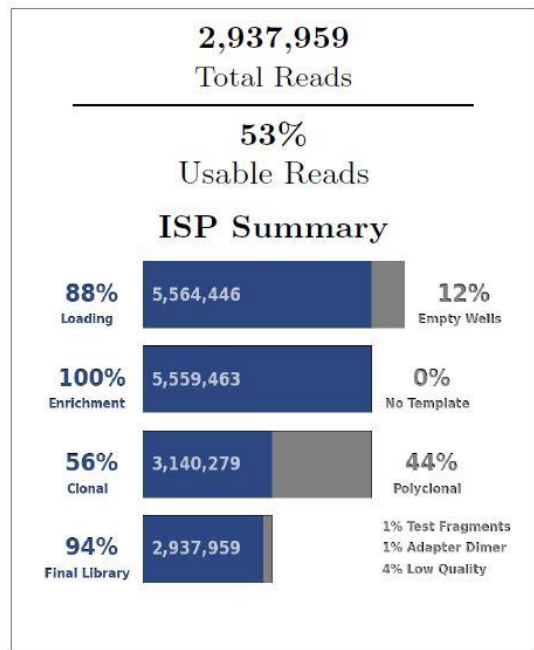
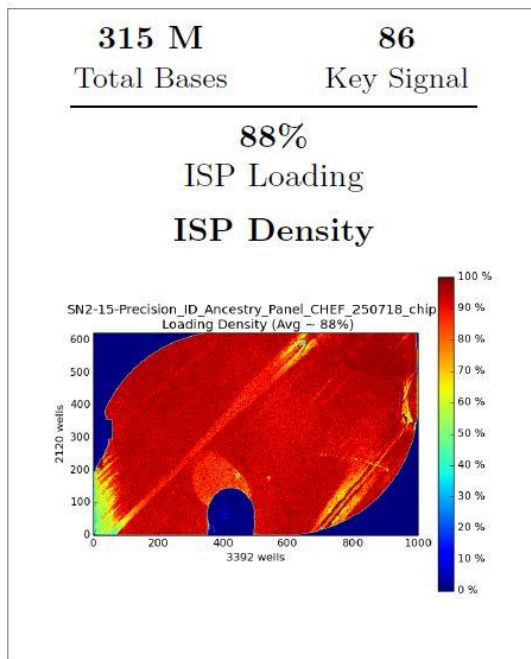
Table 6.9. Data generated from 48 samples following automated library preparation. Sixteen samples out of 48 were included in both workflow (manual and automation workflow) and are highlighted in grey colour. Yellow cells represent the samples that were amplified using 22 cycles whereas blue cells represent the sample that were amplified for 25 cycles (these are slightly different than the recommended usage for the manual workflow, which advise 21 and 26 cycles).

Sample type	Tube 1	Tube 2	Tube 3	Tube 4	Tube 5	Tube 6	Total
Hair	0	0	0	0	1	0	1
Cells	2	2	0	1	0	2	7
Semen	0	0	0	1	0	1	2
Bone	1	2	0	0	0	0	2
Teeth	0	1	0	0	0	0	1
Blood	2	0	0	2	5	0	9
saliva	2	1	8	1	1	2	15
Touch	0	2	0	3	1	3	9
Tissue	1	0	0	0	0	0	2
Total	8	8	8	8	8	8	48
Chip#	Chip#1		Chip#2				

6.4.1 Quality of the sequencing runs

The Summary Reports were used to study and evaluate the chip quality through ISP Density, ISP Summary, and Read Length was used to evaluate each run. First, the quality of chip loading was evaluated using Ion Sphere™ Particles (ISPs) for each run. For this workflow, two 316 v2 Chips were used to sequence 48 samples with 24 samples in each chip. The chip loading densities were 88% for Chip#1 and 82% for chip#2. The ISP summary, which explained the quality of the ISPs that were loaded onto the two chips (Figure 6.3). This was followed by the read length histogram which described the distribution of the read length of HID Ancestry panel. The mean read lengths for all the 2 chips were 107 bp, 108 bp, respectively of the read length as shown in the third part as a histogram.

Chip#1



Chip#2

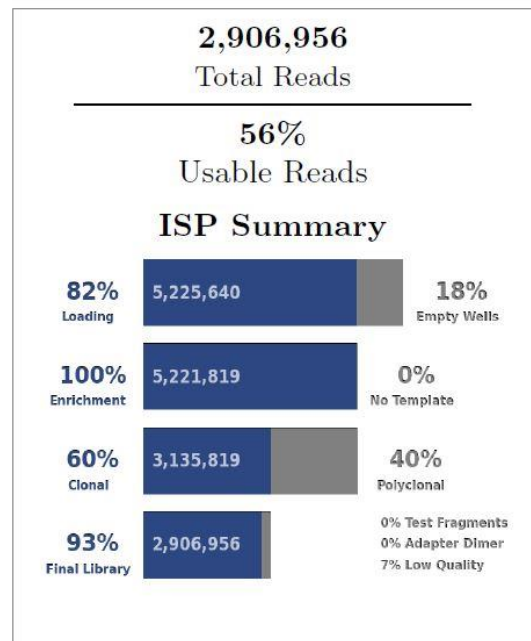
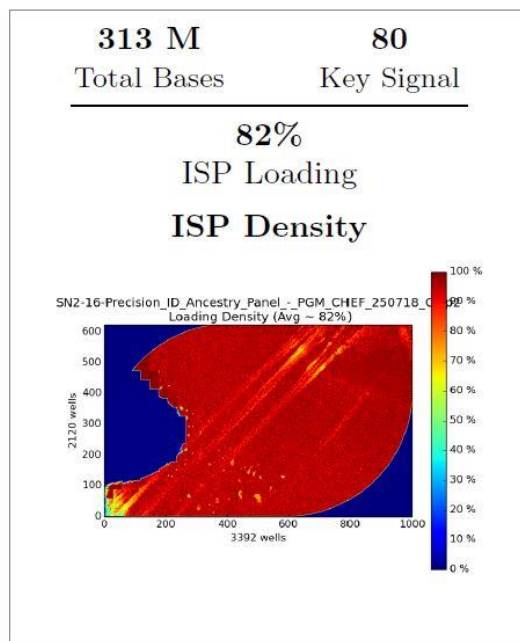


Figure 6.3. Diagrams showing the ISP density percentage and ISP summary for Chip#1 and Chip#2 which was taken from summary report for each chip.

6.4.2 Sequence Coverage

The average mean of depth was obtained from the coverage analysis report as shown in Table 6.10 for two 316™ Chips. Table 6.11 shows the lowest coverage values that were observed and shown among all samples and also Figure 6.4 shows the maximum and median values for each SNP from all the casework samples, the average median value was 613.62x with range from 77.5x (rs2504853) to 1382.5x (rs1079597).

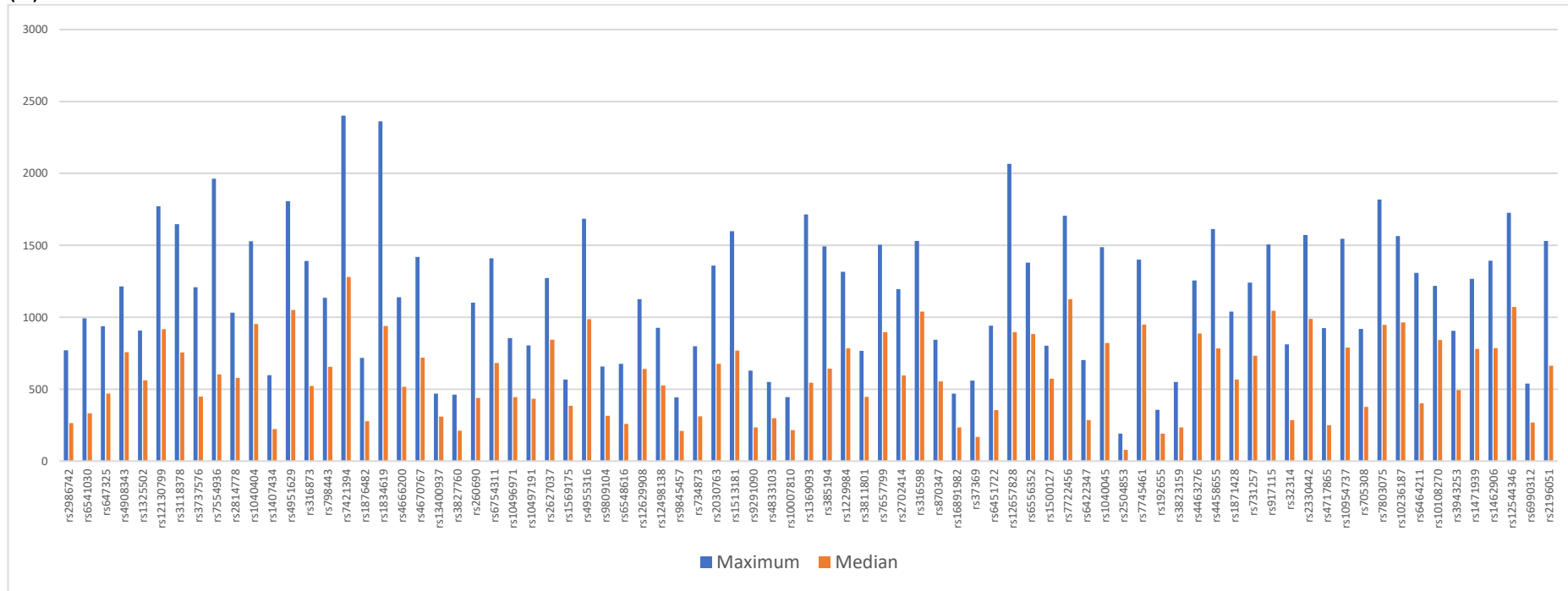
Table 6.10. Data showing the average mean depth for Chip#1 and Chipe#2 used in automation workflow.

Chip#	Average Mean Depth
1	654.38
2	658.90

Table 6.11. Data showing the SNPs with the lowest coverage values.

SNP IDs	Number of samples	Chips#
rs2504853	26	1,2
rs1296819	5	1,2
rs192655	5	1,2
rs10512572	4	1,2
rs2196051	2	1,2
rs1871534	2	1,2
rs2986742	1	1
rs2306040	1	2
rs12498138	1	2

(A)



(B)

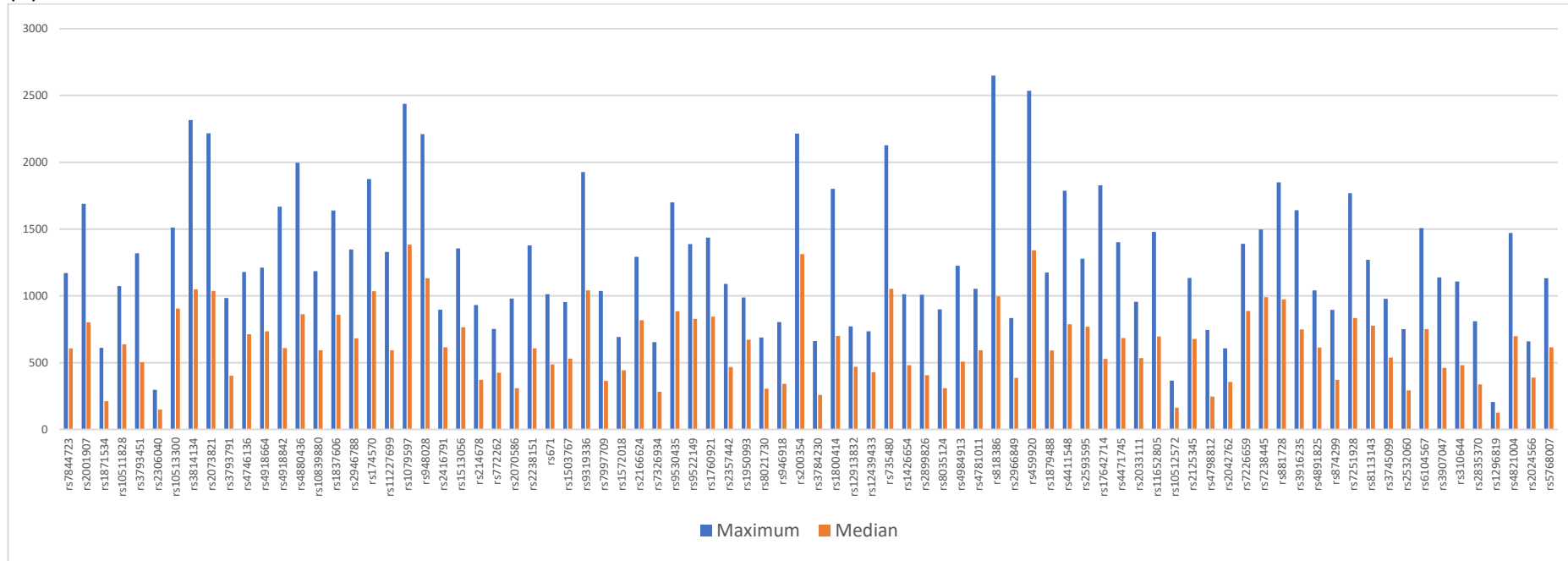


Figure 6.4. Column charts (A) and (B) showing maximum and median of the coverage for each SNP included in the Precision ID Ancestry panel resulted from casework samples (Automation library preparation experiment).

6.4.3 Performance of the Precision ID Ancestry Panel

6.4.3.1 Full SNP profiles

Complete SNP profiles were collected from 29 samples (Table 6.12). Ten samples were already studied with manual library workflow and this was explained in part one of chapter 6 (highlighted in blue colour). Of the 10 samples, 9 of them were case work samples and one sample from the population experiment (Chapter 5). The full profiles were obtained from samples which had DNA quantity of 0.85 ng, 0.32 ng, 0.24 ng, 0.19 ng, 0.14 ng, respectively. The samples used here were from different types of biological material including challenged evidence collected from 2005, 2016, 2017, and 2018.

Table 6.12. Data showing the full SNP profiles which were seen in 29 samples. Samples in grey were previously examined using the manual workflow.

Sample#-Barcode#-Plate#	Years of Analysis	DNA Concentrations (ng)*	Sample Types	STR results
S1-BC1-Ch.1	2005	10.34	Cells collected from teeth cleaning twig "Miswak"	PPlus, Re-amp-GF/FP
S2-BC2-Ch.1	2005	76.44	Tissue	PPlus/FP
S3-BC3-Ch.1	2015	17.02	Bone	IDP, Re-amp-GF/FP
S5-BC5-Ch.1	2017	1.3	Cells from tooth stick	IDP/FP
S6-BC6-Ch.1	2017	4.28	Blood stain	GF/FP
S7-BC7-Ch.1	2016	27.90	Saliva Stain	IDP/FP
S8-BC8-Ch.1	2017	37.10	Blood stain	GF/FP
S10-BC10-Ch.1	2018	0.32	Cells collected from tissue paper	GF/FP

Sample#- Barcode#- Plate#	Years of Analysis	DNA Concentrations (ng)*	Sample Types	STR results
S11-BC11-Ch.1	2017	0.14	Touch. Swab from a gun	GF/FP
S17-BC17-Ch.1	2017	13.72	Saliva	IDP/FP
S18-BC18-Ch.1	2016	49.34	Saliva	IDP/FP
S19-BC19-Ch.1	2016	76.88	Saliva	IDP/FP
S20-BC20-Ch.1	2016	40.17	Saliva	IDP/FP
S21-BC21-Ch.1	2016	30.62	Saliva	IDP/FP
S22-BC22-Ch.1	2016	8.22	Saliva	IDP/FP
S23-BC23-Ch.1	2016	29.83	Saliva	IDP/FP
S24-BC24-Ch.1	2016	4.89	Saliva	NT
S9-BC1-Ch.2	2018	2.22	Blood stain	GF/FP
S10-BC2-Ch.2	2018	50.15	Blood stain	GF/FP
S12-BC4-Ch.2	2018	18.58	Saliva stain	GF/FP
S13-BC5-Ch.2	2018	16.39	Blood stain	GF/FP
S14-BC6-Ch.2	2018	1.25	Blood stain	GF/FP
S21-BC13-Ch.2	2017	0.85	Cells collected from bed blanket	GF/FP
S22-BC14-Ch.2	2017	2.10	Saliva from a Cigarette	GF/FP

Sample#- Barcode#- Plate#	Years of Analysis	DNA Concentrations (ng)*	Sample Types	STR results
S2-BC26-Ch.2	2018	6.01	Blood stain	GF/FP
S3-BC27-Ch.2	2018	0.2	Blood stain	GF/FP
S4-BC28-Ch.2	2018	0.19	Touch-swab from knife blade-handle	GF/FP
S8-BC32-Ch.2	2018	9.9	Semen stain	GF/FP

* All DNA sample concentration was adjusted to 1 ng and volumes based on quantity of DNA. FP- Full profile, PP- Partial profile, NT- Not tested, GF- Global Filer, IDP- Identifiler Plus and PPlus- Profiler plus. Re-amp- sample reamplified; according to the extract volume available in the sample tube, some samples were reamplified with GlobalFiler.

6.4.3.2 Partial SNP profiles

Partial profiles were seen in 18 casework samples with different levels of dropout (Table 6.13). Six samples generated partial profiles with only one dropout SNP, and the DNA quantification values for the samples were: 10.21 ng, 0.7 ng, 0.34 ng, 0.52 ng, 46.12 ng and 0.04 ng. Two SNPs were not genotyped in one sample with a DNA concentration 0.61 ng whereas three samples generated the profile from 161 SNPs out of 165. Seven samples showed partial profiles with 6 (one sample), 7 (two samples), 8 (one sample), 9 (one sample), 11 (one sample) and 15 (one sample) drop-outs. The tooth sample with a DNA input of 0.01 ng displayed the highest dropout with only 69 SNPs profiled (96 SNPs with no data).

Table 6.13. Data showing the partial SNP profiles were seen in 18 samples, rows highlighted in blue were already studied with manual library workflow.

Samples#- Barcodes#- Plates#	Year of Analysis	DNA Concentration (ng)*	Sample Type	#SNP missed the Genotype	Percentage of profile recovery	STR results
S4-BC4-Ch.1	2016	10.21	Saliva	1	99.39	IDP/FP
S9-BC9-Ch.1	2016	0.13	Touch swab from handle scissors	11	93.33	GF/FP
S12-BC12-Ch.1	2016	0.7	saliva-chewing gum	1	99.39	IDP, Re- amp- GF/FP
S13-BC13-Ch.1	2018	0.61	Cells collected from traditional male headdress.	2	98.79	GF/FP
S14-BC14-Ch.1	2017	0.01^	Teeth	96	41.82	GF/PP
S15-BC16-Ch.1	2016	0.01^	Bone	7	95.76	GF/FP
S1-BC25-Ch.2	2018	0.34	Touch. Swab from screw	1	99.39	GF/FP
S5-BC29-Ch.2	2018	0.19	swab from knife blade-handle	9	94.55	GF/FP
S6-BC30-Ch.2	2018	0.52	Cells. scratch collected from inside hat	1	99.39	GF/FP
S7-BC31-Ch.2	2018	6.3	Saliva from Cigarette	4	97.58	GF/FP
S11-BC3-Ch.2	2018	0.34	Touch. Swab from screw	4	97.58	GF/FP
S15-BC7-Ch.2	2018	46.12	Blood stain	1	99.39	GF/FP
S17-BC9-Ch.2	2018	0.04^	Touch. Swab from car door handle	1	99.39	GF/FP
S18-BC10-Ch.2	2017	0.04^	Touch. Swab from cables	7	95.76	GF/FP
S19-BC11-Ch.2	2017	0.04^	Cells collected from a plastic bottle	4	97.58	GF/FP
S20-BC12-Ch.2	2018	0.25	Semen stain	6	96.36	GF/FP
S23-BC15-Ch.2	2017	0.19	Saliva from Cigarette	8	95.15	GF/FP
S24-BC16-Ch.2	2018	0.10	Touch. Swab from cables	15	90.91	GF/FP

* All DNA sample concentrations were adjusted to 1 ng and volumes based on quantity of DNA. ^ The maximum volume of extracts (15 µl) was added in case of some samples yielded concentrations of DNA below that value. FP- Full profile, PP- Partial profile, GF- Global Filer and IDP- Identifiler Plus. Re-amp- sample reamplified; according to the

extract volume available in the sample tube, some samples were reamplified with GlobalFiler.

6.4.3.3 No SNP profiles

In this experiment, only one sample failed to generate a profile. This was a bone powder sample from 2016 with 0.15 ng of input DNA which was amplified with Global Filer and provided a partial profile.

Part three: Ancestry Inference of the Precision ID Ancestry panel

6.5 Results

6.5.1 Ancestry prediction with the default HID SNP Genotyper

HID SNP Genotyper is a plugin for Torrent Suite software (TFS) which provides Admixture prediction and Population likelihood results based on genotype produced for all SNP markers. The reports were exported as pdf files from summary reports. The Admixture prediction and Population likelihood were obtained for the casework samples in manual library preparation workflow. Most of the casework samples (84 samples) were predicted to have Asian origin and they included Southwest Asia, South Asia and East Asia ancestry with 76.68% (Figure 6.5). The samples were reported to have as Kuwaiti, Keralite, Palestinian, and Hazara in origin. Some samples indicated to have as Jews, Yemenite, Druze and Negroid Makrani in origin. Nine samples were predicted to have African ancestry with 8.11% and fewer samples (6 samples) showed to have as Europe ancestry with 5.4%. Only one sample was assigned to have Oceania ancestry (Table 6.14).

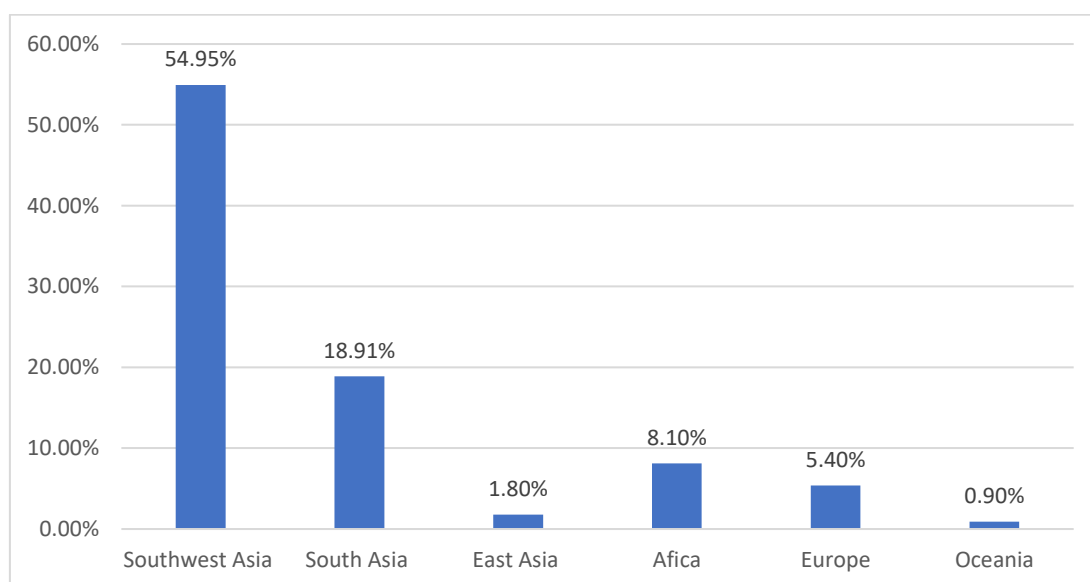


Figure 6.5. Bar charts showing the geographic distribution for 100 casework samples. Data were expressed as a percentage.

Table 6.14. Data showing the Population Likelihood results from the most likely population of origin of the 100 DNA profiles.

Population Names	Total number of individuals
Kuwaiti	23
Keralite	16
Palestinian	10
Hazara	8
Jews, Yemenite	6
Druze	6
Negroid Makrani	4
Somali	4
Kachari	3
Pashtun	3
Mohanna	3
Samaritans	2
Jews, Ethiopian	2
Hakka	1
Khanty	1
Lao Loum	1
Samoans	1
Adygei	1
Jews, Sephardic	1
Russians	1
Greek	1
Hausa	1
Ibo	1
African Americans	1
Total	100

HID SNP Genotyper Plugin v5.2.2 was used for the secondary sequence analysis. The Admixture Prediction and Population Likelihood results were reviewed.

Most of the samples (40 samples) in in automation library preparation workflow experiment were predicted to have Asian origin from Southwest Asia, South Asia and East Asia ancestry with 83.30%. The populations recorded were Negroid Makrani, Keralite, Kuwaiti Jews Yemenite and Palestinian. Four samples were predicted to have African ancestry with 8.30% and 2 samples were showed to have as European ancestry with 4.16%. Only one sample was assigned Oceanian individual (Table 6.15 and Figure 6.6).

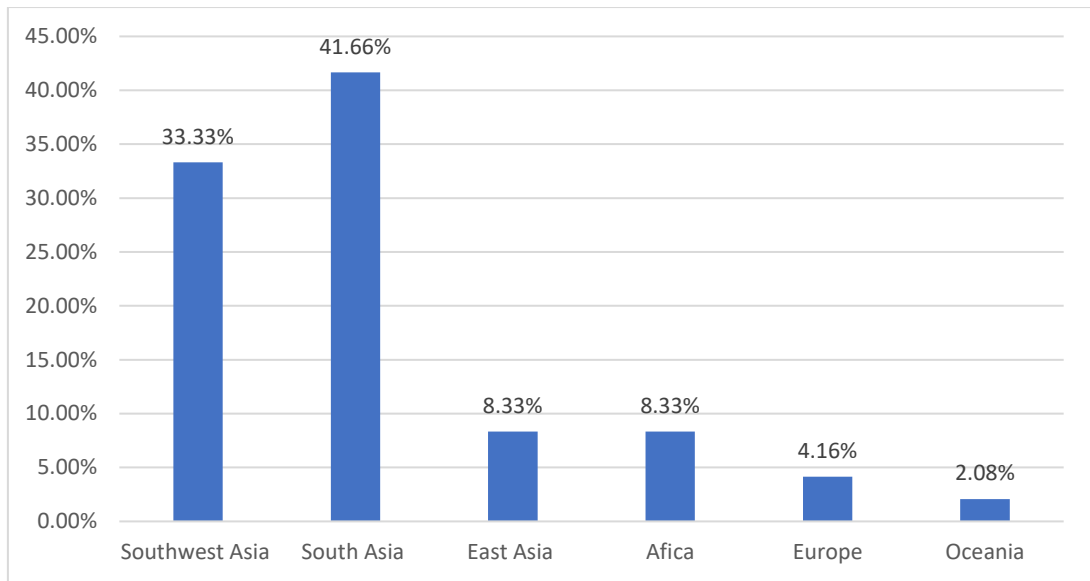


Figure 6.6. Bar charts showing the geographic distribution for the 47 samples processed with automation workflow. Data expressed as a percentage.

Table 6.15. Data showing the Population Likelihood results. The most likely population of origin of the 47 DNA profiles.

Population Names	Total number of individuals
Negroid Makrani	7
Keralite	6
Kuwaiti	5
Jews, Yemenite	5
Palestinian	4
Hazara	3
Lao Loum	3
Jews, Ethiopian	3
Kachari	2
Pashtun	2
Druze	2
Cambodians	1
Samoans	1
Adygei	1
Greek	1
Somali	1
Total	47 samples (one sample with No profile)

6.5.2 Ancestry prediction with the customized HID SNP Genotyper with Qatar population Data

A data set including the allele frequencies of Qatari samples was prepared using PowerStat Microsoft Excel Workbook (Tereba 1999). It was sent to ThermoFisher Scientific to design a custom HID SNP Genotyper plug-in for Qatari population. The 101 samples analysed in manual library workflow and the 48 samples used in the automatic library system were re-analysed after Qatar Data was added to HID SNP Genotyper plugin in v5.2.2.

The re-analysis of 100 samples which generated full and partial profiles from the manual library workflow are listed in (Table 6.16). Thirteen samples were assigned to have the highest likelihood to Qatar population (highlighted in pink). The 13 samples included 12 samples which were predicted to have as Southwest Asia and the rest were Jews Yemenite (4), Palestinian (4), Kuwaiti (2) and Druze (2) in origin and 1 sample was predicted to be as Greek in origin (highlighted in grey).

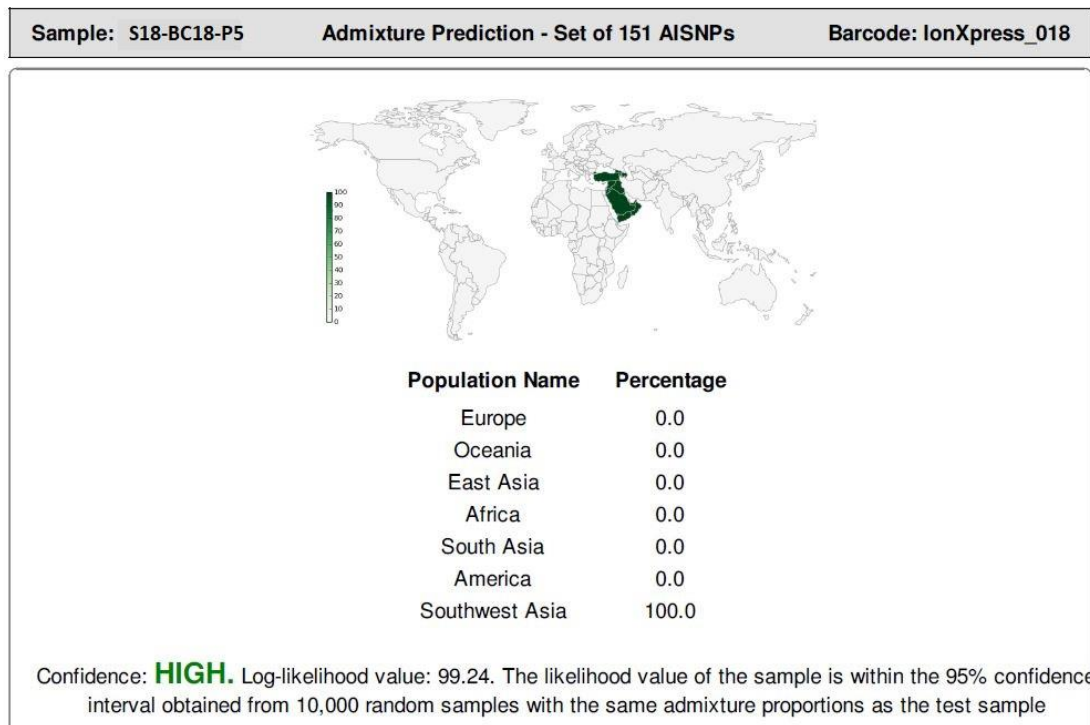
Table 6.16. Data showing the re-analysis with Qatari Population Likelihood results with the most likely population of origin of the 100 DNA profiles (Manual Workflow).

Total number of individuals Without Qatar Data		Total number of individuals customized HID SNP Genotyper	
Kuwaiti	23	Qatar	13
Keralite	16	Kuwaiti	21
Palestinian	10	Keralite	16
Hazara	8	Palestinian	6
Jews, Yemenite	6	Hazara	8
Druze	6	Jews, Yemenite	2
Negroid Makrani	4	Druze	4
Somali	4	Negroid Makrani	4
Kachari	3	Somali	4
Pashtun	3	Kachari	3
Mohanna	3	Pashtun	2
Samaritans	2	Mohanna	3
Jews, Ethiopian	2	Samaritans	2
Hakka	1	Jews, Ethiopian	2
Khanty	1	Hakka	1
Lao Loum	1	Khanty	1
Samoans	1	Lao Loum	1
Adygei	1	Samoans	1
Jews, Sephardic	1	Adygei	1
Russians	1	Jews, Sephardic	1
Greek	1	Russians	1
Hausa	1	Hausa	1
Ibo	1	Ibo	1
African Americans	1	African Americans	1
Total	100	Total	100

The results in Table 6.17 were obtained from the samples that were used in automation workflow. The data show that the 47 samples were successfully processed generating either full or partial SNP. Five samples out of 40 were reported to have an Asian (Southwest Asia) and were assigned to Kuwaiti, Jews Yemenite, Palestinian and Druze (highlighted in grey). As such, those samples were observed to have the highest value for Qatar (highlighted in pink) (Figure 6.7).

Table 6.17. Data showing the re-analysis with Qatar Population Likelihood results; the most likely population of origin of the 47 DNA profiles (Automation Workflow).

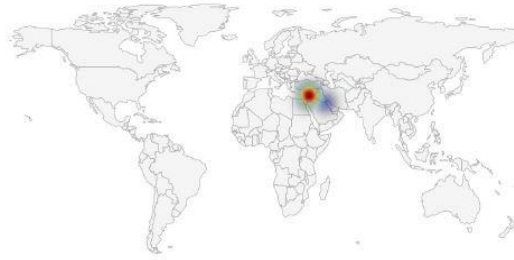
Total number of individuals Without Qatar Data		Total number of individuals customized HID SNP Genotyper	
		Qatar	5
Negroid Makrani	7	Negroid Makrani	7
Keralite	6	Keralite	6
Kuwaiti	5	Kuwaiti	4
Jews, Yemenite	5	Jews, Yemenite	3
Palestinian	4	Palestinian	3
Hazara	3	Hazara	3
Lao Loum	3	Lao Loum	3
Jews, Ethiopian	3	Jews, Ethiopian	3
Kachari	2	Kachari	2
Pashtun	2	Pashtun	2
Druze	2	Druze	1
Cambodians	1	Cambodians	1
Samoans	1	Samoans	1
Adygei	1	Adygei	1
Greek	1	Greek	1
Somali	1	Somali	1
Total 47 samples (Full and Partial SNP profiles) (one sample with No profile)			



Sample: S18-BC18-P5

Population Likelihoods - Set of 151 AISNPs

Barcode: IonXpress_018

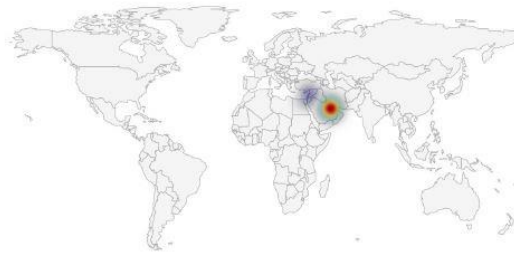


Population Name	Geo Region	Likelihood
Druze	Asia	6.17E-44
Palestinian	Asia	2.19E-44
Kuwaiti	Asia	9.43E-45
Jews, Yemenite	Asia	3.25E-46
Sardinian	Europe	2.54E-46
Jews, Ashkenazi	Europe	4.04E-47

Sample: S18-BC18-P5

Qatar Population Likelihoods

Barcode: IonXpress_018



Population Name	Geo Region	Likelihood
Qatar	Asia	8.76E-43
Druze	Asia	6.17E-44
Palestinian	Asia	2.19E-44
Kuwaiti	Asia	9.43E-45
Jews, Yemenite	Asia	3.25E-46
Sardinian	Europe	2.54E-46
Jews, Ashkenazi	Europe	4.04E-47

Figure 6.7. Maps and data showing the Admixture Prediction, Population Likelihood and Qatar Population Likelihood result of a typical example number S18-BC18-P5. It was a saliva stain from a cigarette butt with a DNA input 1.9 ng.

6.6 Discussion

This chapter of the study evaluated the Precision ID Ancestry Panel from Thermo Fisher Scientific and it is considered as the second evaluation experiment of MPS workflow using Precision ID Ancestry Panel and the Ion Torrent Personal Genome Machine (PGM). Several types of biological evidence that varied from routine stains in some samples had different levels of challenge and as such they were used in the study. The samples were collected from real cases between 2005 and 2018 and the DNA was extracted by different extraction methods that are used in routine casework analysis in the lab. A total of 143 casework samples from 76 real cases were selected and they represented both fully and partially genotyped samples. Samples had been previously genotyped with one or more of Identifiler[®], Identifiler[®] Plus, MiniFiler[™], and GlobalFiler[®]. Some samples were amplified with Yfiler[®] and Yfiler Plus[®] (Applied Biosystems[™]).

In this study, both manual and automated library preparation workflows were tested. For the manual preparation, a total of 111 casework samples were used and the Ion Chef instrument was used for template preparation and Chip loading and 48 samples were used on automation library preparation using the Ion Chef.

The results collected from both experiments showed the ability of the panel to be applied for most casework samples. Full and partial profiles were successfully generated from most of the casework samples studied in both workflows. Increased PCR cycling conditions lead to increased panel sensitivity which helped to obtain full profiles from weak samples as seen down to 0.12 ng with high confidence level. This was a swab from a cup.

The Manufacturer's user guide recommends extending the PCR cycles for sample input DNA less than 1 ng and that was applied in both workflows. In Manual workflow, the amplification cycles used were 21 and PCR cycle number were increased to 5 more cycles (i.e. 26 cycles) as recommended. Whereas in automation workflow, the amplification was for 22 and 25 cycles. Both experiments performed well as the tested samples were real casework samples. From the results generated in this study and from the population experiment, it would be an introduction to start to design threshold through using coverage values, heterozygous balance and noise level. The

threshold will include minimum and maximum coverage value and heterozygous balance.

Between the two workflows methods, there were some points observed as genotype quality. Six samples had shown full profiles in both experiments (highlighted in purple). In automation workflow, complete SNP profiles were seen in 3 samples, whereas the samples in the manual experiment were partial (highlighted in green). Also, incomplete SNP profiles were observed in 3 samples with varying amount of drop-out (highlighted in blue). An improvement was seen in two samples that did not give any SNP profiles in the manual workflow. One of the samples showed a full profile and the other sample was a partial profile with one missing SNP (highlighted in pink). One sample that failed to generate a profile in the automation workflow had a partial profile with 102 SNPs dropping out (Table 6.18).

Table 6.18. Data showing the profile quality of the 16 samples studied in both workflow: Manual and Automation library workflow.

Sample Information		Manual Workflow		Automated Workflow	
Sample Name	DNA Conc.	Profile Quality	# of SNPs dropped	Profile Quality	# of SNPs dropped
S1-BC1-Ch.1	10.34 ng	Full	0	Full	0
S2-BC2-Ch.1	76.44 ng	P	2	Full	0
S3-BC3-Ch.1	17.02 ng	Full	0	Full	0
S4-BC4-Ch.1	10.21 ng	P	1	P	1
S5-BC5-Ch.1	1.3 ng	Full	0	Full	0
S6-BC6-Ch.1	4.28 ng	Full	0	Full	0
S7-BC7-Ch.1	27.90 ng	Full	0	Full	0
S8-BC8-Ch.1	37.10 ng	Full	0	Full	0
S9-BC9-Ch.1	0.13 ng	P	10	P	11
S10-BC10-Ch.1	0.32 ng	P	7	Full	0
S11-BC11-Ch.1	0.14 ng	P	39	Full	0
S12-BC12-Ch.1	0.7 ng	NR	NR	P	1
S13-BC13-Ch.1	0.61 ng	P	9	P	2
S14-BC14-Ch.1	0.01 ng	P	54	P	96
S15-BC16-Ch.1	<1ng	P	8	P	5
S16-BC16-Ch.1	<1ng	P	102	NR	NR
S24-BC24-Ch.1	4.89 ng	NR	NR	Full	0

Coverage distribution across all the SNPs varied as mentioned above from highest to lowest. The locus rs1296819 received the lowest coverage in both experiments with 20 samples in the manual experiment and 5 samples in the automated workflow. This was also observed in 22 samples of Qatari population as detailed in Chapter 5 (Table 6.19

and Table 6.20). This SNP rs1296819 was also reported to receive lowest coverages studies (Al-Asfi et al., 2018; Pereira et al., 2017; Santangelo et al., 2017).

Table 6.19. Data showing the SNPs with the lowest coverage with the two workflows.

SNP IDs	Manual/number of samples	Automation/ number of samples
rs1296819	20	5
rs192655	9	5
rs2196051	7	2
rs2986742	8	1
rs2504853	1	26
rs12498138	5	1

Table 6.20. Results showing the SNPs with the lowest coverage values seen in two experiments.

SNP/IDs	Population samples experiment	Forensic casework samples experiment
rs1296819	22	25
rs260690	55	13
rs1950993	63	5
rs192655	4	14
rs2504853	1	27

The Ion Chef was used for the template preparation of AmpliSeq libraries and chip loading. Ion Chef provides an automation system for template preparation including the emulsion PCR, enrichment and two chips reading for sequencing on Ion PGM. It reduced the time for the preparation and pipetting working. Consistent chip loading helps to have very good loading percentages. Low values of ISP density can affect the results interpretation such as observing partial genotypes and alleles drop outs (Mogensen et al., 2015). Similar results were obtained with the Ion chef instrument for the automation library workflow. The automation workflow only was done by using 8 samples in each run, but it was shown as a noticeable difference in the time used in the manual workflow; which would extend to 2 days. The automated workflows were found to offer a sizable reduction in hands-on time compared to the manual workflow (van der Heijden et al., 2017).

Ancestry inference was assessed using the HID-SNP Genotyper plugin which carried out two forms of predicting, including Admixture Prediction and Population Likelihood. The newly available ancestry data have the potential to provide information on

unidentified DNA donors that can feed into and inform a police investigation (Hollard et al., 2017). As well as aiding in criminal investigations, this type of information may be useful in cases of missing persons where skeletal remains can be analysed. This is of special reference in the absence of a hit on a national DNA database which may provide some relevant information (Hollard et al., 2017).

In this study, attempts have been made to apply Precision ID Ancestry panel to casework samples. A study by Al-Asfi et al. (2018) demonstrated high sensitivity and reproducibility of the panel for use in forensic casework with high quality single source DNA samples. Similarly, the panel was studied by Hollard et al. (2017) in a real forensic case, where a carbonized body was found with no physical description or personal belongings that could be used for identification. The DNA profile did not result in a match with the database, but by using HID-Ion AmpliSeq Ancestry Panel, a probable geographical origin could be determined (Bruijns et al., 2018; Hollard et al., 2017).

Qatar, is one of the smallest States on the Arabian Peninsula, but has a large flow of transnational labours (Gardner et al., 2014). The State of Qatar has initiated numerous infrastructure and development projects that vastly exceed the indigenous labour supply. The influx of temporary guest workers is primarily from West and South Asia (Omberg et al., 2012). This urban and the great openness helped the State to be a destination for many nationalities. Currently, more than 80 countries are eligible for visa-free entry into the State of Qatar, with variable lengths of stay (Ministry of Interior of Qatar, 2018).

As mentioned in Chapter 5, the collection of 300 Qatari samples from the eight municipalities was helped to see geographical distribution and the diversity of Qatar population using the HID Ancestry Panel. Moreover, it was equally important to mention the people who came to Qatar across the Qatari mainland and across the Eastern coast of the Arabian Gulf (Obaidan, 1982). In addition, the collection of 143 unknown samples included predictions for population which were not seen with the 300 Qatari population samples. They included people from East and South Asia as Hazara, Kachari, Hakka, Lao Loum and Khanty as well as populations from Africa such as as Ibo, Hausa and also from Europe as Russians and Samoans from Oceania.

Table 6.21 Data showing the predicted populations for the Qatari population and casework samples studied with Precision ID Ancestry Panel. (Grey cells represent the population seen in both experiments, yellow only in population samples and green cells only in casework samples).

Population samples	Casework samples
Qatar	Qatar
Palestinian	Palestinian
Kuwaiti	Kuwaiti
Druze	Druze
Jews, Yemenite	Jews, Yemenite
Pashtun	Pashtun
Negroid Makrani	Negroid Makrani
Jews, Sephardic	Jews, Sephardic
African Americans	African Americans
Jews, Ethiopian	Jews, Ethiopian
Somali	Somali
Keralite	Keralite
Adygei	Adygei
Greeks	Greeks
Mohanna	Mohanna
Samaritans	Samaritans
Sardinian	Hazara
Jews, Ashkenazi	Lao Loum
Sandawe	Russians
	Hausa
	Ibo
	Cambodians
	Samoans
	Kachari
	Khanty
	Hakka

This result was expected due to the different nationalities of expatriates residing in the country, whether for work or for visit, from different regions.

Ancestry estimation, using the Precision ID Ancestry pane, can calculate the likelihood of the genotype in each reference population. Likelihood values cannot be interpreted as absolute numbers. As described by Rajeevan et al., (2012) "While the population with the highest likelihood is the most likely origin of the input genotype, it is not necessarily the correct origin and those with similar, albeit lower likelihoods, cannot be excluded'. It is therefore important not to focus only on the population with the highest likelihood but also on all populations with a comparable likelihood (Hollard et al., 2017; Rajeevan et al., 2012).

Admixture assessment and interpretation are complex processes that requires the presence of the origin of populations among the reference populations. Moreover, admixture, especially in a geographical region that is between major populations, will inevitably be complex. Southwest Asia, as an interface between Europe, Africa and South Asia, makes it difficult to identify if admixture is present (Hollard et al., 2017). Determining the origin of a sample is complex, especially in the presence of patterns of admixture and with limited reference data. Assignment for individuals from multiple populations of origin is a more complex task than for individuals from a single population of origin (Al-Asfi et al., 2018). Studying and adding reference population from different regions that are not in the data set could improve the inference process. Considering the information collected from both experiments and the addition of Qatari population dataset to the plug-in, it was useful to sort the results. When the population samples were reanalysed with the plug-in included the Qatari samples, some samples were changed to be predicted as Qatari and the previous highest likelihood predicted went down. However, there were some samples which remained unchanged and the Qatari population likelihood went to the second or third order and lower.

How these data could be incorporated into casework is challenging. In both the population and the casework samples the vast majority of samples fall into the Southwest Asian or South Asian category and therefore would be of limited assistance to investigators. There may be cases where the suspected perpetrators do not come from South Asia/Southwest Asia, where the employment of this technology could be helpful from the point of an investigation.

6.7 Conclusion

In conclusion, the results collected from the current study have demonstrated that the use of this MPS panel is a promising tool in the forensic field as representing the benefit of the combining approach with the routine STR test results. Countries which are located in Southwest Asia makes the prediction process in some samples difficult especially since it is the crossroads of migrations. As such, it would be more useful and beneficial when the panel applied to the different nationalities living in Qatar; to enhance the predictive qualities more information from worldwide populations are needed.

CHAPTER: 7 General Discussion, conclusion, limitation and future work

7.1 General Discussion

Modern DNA sequencing technologies, in particular, the massively parallel sequencing (MPS) have the potential to revolutionize forensic genetics. For the first time, the technology has enabled the simultaneous profiling of hundreds of SNPs (Lee et al., 2018; Pereira et al., 2017) and opens up the potential to investigate markers for ancestry and phenotype as well as identity (Bruijns et al., 2018).

The present study consisted of 4 main series of experiments and 289 SNP markers were applied to the Qatari population samples and to selected casework samples (known and unknown). The sequencing performance, forensic parameters, ancestry inference and Y-haplogroup distributions were investigated.

Both panels are designed to amplify the samples as little as 1 ng of input DNA and optimized for challenged samples. As shown in both chapters 3 and 5, the sensitivity experiment showed that the full SNP profile was seen down to 0.25 ng (in triplicates). Also, lower template of 0.05 ng yielded full profiles in some of the replicates whereas the lowest input 0.01 ng showed a large number of drop-out loci.

One of the most important goals of the thesis was to study the suitability of the panels for use in the analysis of forensic casework samples. A variety of real forensic casework samples collected between 2005 and 2018 were profiled from routine to challenging samples. Increased PCR cycling conditions lead to increased panel sensitivity which helped to collect full profiles from weak samples down to 0.12 ng in the Ancestry experiment. This was a swab from a cup analysed in 2005 with Profiler Plus. With the Identity panel complete profiles were observed down to 0.06 ng; the sample was soft tissue from femur, which had been analysed previously with Identifiler Plus.

Published studies with the Precision ID Identity Panel are limited. Guo et al. (2016) reported that the HID ID Identity panel had the ability to be worked with routine and different casework samples and the full SNP genotype was seen with 100 pg of DNA input. Similarly, Meiklejohn and Robertson (2017) found reproducible profiles could be generated with ≥ 0.2 ng of template DNA.

The MPS workflows are complex and include numerous pipetting steps for the transfer and mixing of reagents. During the library preparation several reagents used some have high viscosity and the volumes used are small. As such, it must be handled with skill. In this study, Ion Chef was used in both for template preparation of Ampliseq libraries and chip loading and then was also used to automate the library preparation for 48 samples. Ion Chef provides an automation system for template preparation including the emulsion PCR enrichment. Automation of the library building process significantly reduces the hands-on time in the laboratory, decreases the risk of pipetting errors and reduces the workload of the laboratory operator and also reduces the chance of contamination. It is a conclusion from the experiment that the MPS automated workflow (automated library preparation, automated template preparation and chip loading) is the preferable option when/if the forensic lab starts implementing MPS.

7.2 Combined results of the Precision ID Identity panel and Precision ID Ancestry Panel

The study is a typical example of such combining MPS results. The approach was to collect the data from markers applied to the samples whether they were in population samples experiments or in casework experiments.

7.2.1 Population samples results

The HID SNP Genotyper Plugin calculates the Random Match Probability (RMP) as the product of genotype frequencies for each locus. The plugin calculated the RMP for five population groups and they included Africa, Europe, America, East Asia and South Asia based on the 85 unlinked identity SNPs of the 90 SNPs in the panel using allele frequencies obtained from the 1000 Genomes data set. The RMP results for all the 105 unrelated Qatari individuals were listed with results collected from the Precision ID Ancestry Panel experiment detailed in Chapter 5. The results from the Y-haplogroups were predicted and they were based on 34 Y-SNPs, including in the Precision ID Identity panel from the 84 Qatari samples was combined with the results. The 60 individuals assigned to J haplo-group were reported as most likely Qatari, Palestinian, Jews Yemenite, Kuwaiti, Pashtun, Jews Ethiopian, Adygei and Jews Sephardic, Greeks and Pashtun (Appendix 15) (Figure 7.1).

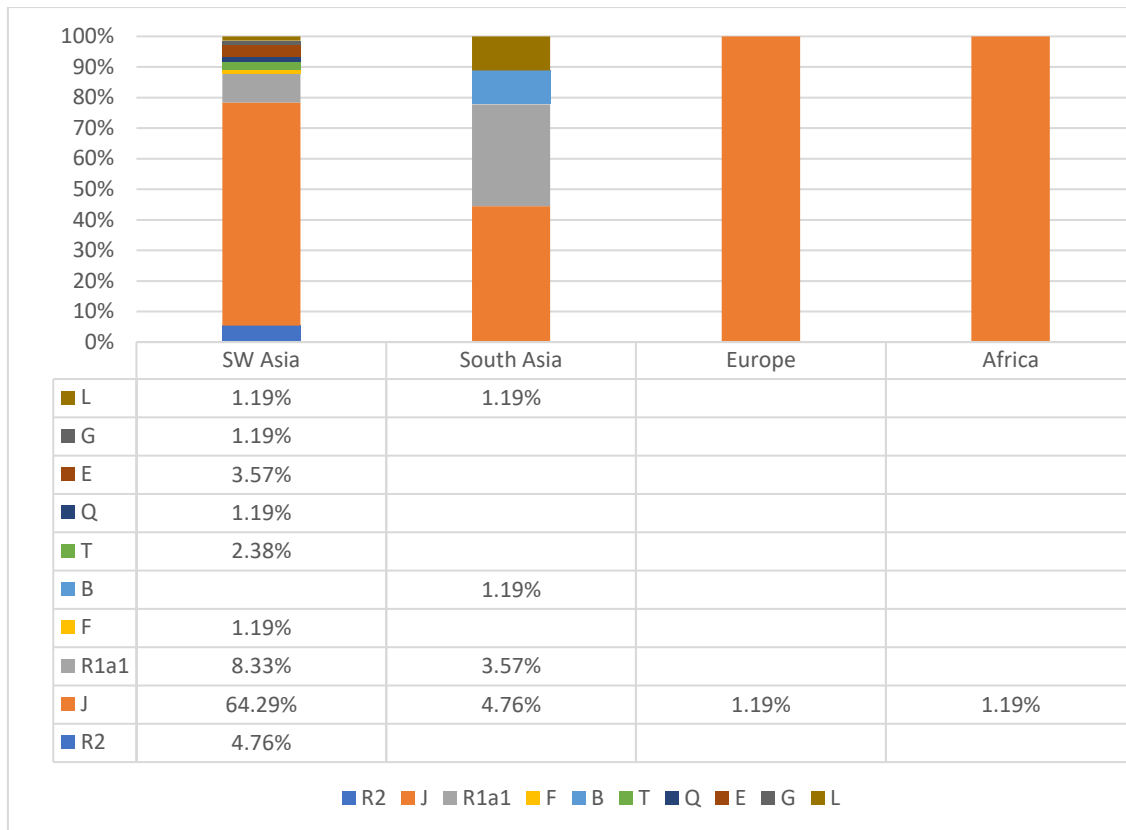


Figure 7.1. Bar charts and tabular data showing the Y haplogroup frequency based on 34 -Y-SNPs included in Precision ID Identity with Geo-region as predicted from Precision ID Ancestry panel.

7.2.2 Forensic samples results

From the forensic results, a table was prepared including the Y haplo-groups predictions of 60 casework samples (as explained in Chapter 4) of which 55 samples with the population likelihood results from experiment described in Chapter 6. They were combined with the RMP results for all the 60 sample (Appendix 16).

The haplo-group of Y chromosomal SNP combination is a valuable marker to determine the haplo-group to which each Y chromosome belongs (Ochiai et al., 2016). The next-generation sequencing (NGS) assay has been applied for widely diverse comprehensive genetic tests.

Of 84 unrelated males from Qatar population, the most common haplo-group was J-S35 (71.43%), followed by R1a1-SRY10831.2 (10.71%), R2-M479 (4.76%), and finally E-P171 (3.57%). The other haplo-groups distributed with lower frequencies, such as T-M184, L-M20 (2.38%) and F-M89, G-M201, Q-M242 (1.19%). The results of this study are consistent with previous investigations that cover the Y-chromosome haplo-group affinity in the Qatari population. Moreover, haplo-group J is the most common component in Qatar and it is distributed in all eight municipalities. The Arabian Y-chromosome can be split into four main groups: J, E, R and T (Al-Zahery et al., 2011; Theyab, 2013). There are still no published studies covering the 34-Y SNPs for the regions neighboring Qatar, where the comparison was made to find the haplo-group distribution in general in Qatar and its neighbors.

Haplo-group J (defined by mutations 12f2a, M304, P209 M304, S6, S34, and S35) is found at the highest frequencies in Middle Eastern and North African populations, Middle Eastern Europe, Central Asia, India, and Pakistan (Karafet et al., 2008; Kaushik et al., 2018). There are two major subclades (J1 and J2), which are defined by mutations M267 and M172. The Southern Arabian Peninsula has been proposed to be the place of origin of haplo-group J (Triki-Fendri et al., 2016). Previous studies showed subclades J1- and J2 are the most common in the Arabian Peninsula with 58% in Saudi Arabia (Abu-Amero et al., 2009), 58.3% in Iraq (Al-Zahery et al., 2011), 66.7% in Qatar, 45.1% in United Arab Emirates (UAE), and 82.2% in Yemen (Cadenas et al., 2008; Theyab, 2013). Another study of 53 male Qataris divided into Bedouin, Persian-South Asian and African, showed that the Arab haplo-group J1 was the dominant haplo-group in the Bedouin Qataris (Rodriguez-Flores et al., 2016). In this study, J-S35 is the most abundant component in Qatar population accounting for 71.43% of samples.

The study also found R1a1 (SRY10831.2) haplo-group in about of (10.71%) of Qatar population. R1a1-SRY10831.2 which is common in East Europe, is also highly represented in the Middle East. In the Kuwaiti population, 9.4% of the samples belonged to Haplogroup R-M173, among which 63% were from the R1a1-SRY10831.2 subhaplogroup (6% of all of the population) (Triki-Fendri et al., 2016), 19.4% in Iraq (Al-Zahery et al., 2011), and 20% in Persians and Tajiks (30%) (Malyarchuk et al., 2013).

The SRY10831.2 is the old European branch of R1a predicted in Klyosov & Rozhanskii (2012) to have originated in Central Asia around 20,000 years before the present (ybp). The bearers of R1a Y chromosome began a migration to the West, through Tibet and over the Himalayas between 20 and 15 thousand years ybp. They arrived in Hindustan no later than 12,000 ybp. They reached across the Iranian Plateau, to East Anatolia and the rest of Asia Minor between 10,000 and 9000 ybp. By ~9000 ybp they were in the Balkans and then spread westward over Europe and to the British Isles (Klyosov & Rozhanskii, 2012; Rozhanskii & Klyosov, 2012).

In the Qatari population, 4.76% % of the males belong to haplo-group R2-M479. Haplogroup R-M207 has two main divisions within its phylogeny, R1-M173 and R2- M479. The R2-M479 branch is found mostly among populations living in South Asia;R2-M124 detected in low frequency in Qatar and Iraq populations (Al-Zahery et al., 2011; Cadenas et al., 2008).

Y-DNA haplo-group E, which has been hypothesized to have arisen in northern Africa, is another Y-chromosome haplo-group identified in the Arabian Peninsula. In the present study haplo-group E (P171) characterizes 3.57% of Qatar population. Haplogroup E is widely distributed across the Middle East, Africa, and the Mediterranean. Recent studies identified the presence of E1b1a, previously known as E3a-M2, which is defined by M2 mutation in Qatar (2.8%) (Cadenas et al., 2008) and also in Oman (Luis et al., 2004), (7.4%),Yemen (3.2%), UAE (5.5%) (Cadenas et al., 2008) and Saudi Arabia 3.8% (Abu-Amero et al., 2009).

The frequencies of both Haplo-groups T (M184) and L (M20) were 2.38% among Qatari individuals in this study. The T haplo-group has been observed at low frequencies in the Middle East, Africa, and Europe (Karafet et al., 2008).

In addition, there are several haplo-groups in the Arabian Peninsula but with low frequency ranging from 0.49% to 3%. The frequencies reported by Cadenas et al. (2008) and this study are shown in Table 7.1.

Table 7.1. Data showing the Y-chromosome haplogroup frequency observed for Qatar collected from the present (N= 84 Qatari) and previous study (N= 72 Qatari).

Haplogroups	Frequency %	References
B2a M150	1.4	(Cadenas et al., 2008)
B2b-M112	1.4	(Cadenas et al., 2008)
B-M181	1.19	Present study
E1b1a7 M191	2.78	(Cadenas et al., 2008)
E1b1b1a* M78	1.4	(Cadenas et al., 2008)
E1b1b1a2* V13	1.4	(Cadenas et al., 2008)
E1b1b1a3* V22	1.4	(Cadenas et al., 2008)
E1b1b1c1* M34	1.4	(Cadenas et al., 2008)
E1b1c M329	1.4	(Cadenas et al., 2008)
E2* M75	1.4	(Cadenas et al., 2008)
E2b* M98	2.78	(Cadenas et al., 2008)
E-P171	3.57	Present study
F- M89+	1.19	Present study
G-M201	1.19	Present study
G2a* P15	2.78	(Cadenas et al., 2008)
H1a* M82	1.4	(Cadenas et al., 2008)
J1* M267	58.3	(Cadenas et al., 2008)
J- S35	71.43	Present study
J2a DYS413	2.8	(Cadenas et al., 2008)
J2a1 M47	1.4	(Cadenas et al., 2008)
J2a2* M67	1.4	(Cadenas et al., 2008)
J2b* M12	1.4	(Cadenas et al., 2008)
J2b2* M241	1.4	(Cadenas et al., 2008)

Haplogroups	Frequency %	References
L1 M76	2.8	(Cadenas et al., 2008)
L-M20	2.38	Present study
R1a1* M198	6.9	(Cadenas et al., 2008)
R1a1- SRY10831.2	10.71	Present study
R1b1b2* M269	1.4	(Cadenas et al., 2008)
R2 M124	1.4	(Cadenas et al., 2008)
R2- M479	4.76	Present study
Q- M242	1.19	Present study
T- M184	2.38	Present study

Y chromosome haplogroups

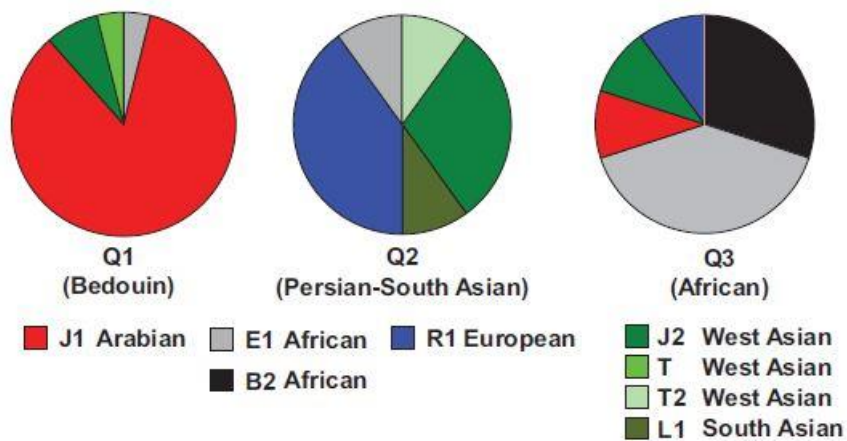


Figure 7.2. Pie charts showing the Rodriguez-Flores et al., (2016) pie charts of the Y haplo-group frequencies among the 53 male Qatari divided into Q1 (Bedouin), Q2 (Persian-south Asian) and Q3 (African).

From the information collected from the Y- haplogroup predictions, it is possible to see the diversity of Qatar population similar to the results obtained from the Precision ID Ancestry Panel, in that the people who came to Qatar across the Qatari mainland and across the Eastern coast of the Arabian Gulf (Obaidan, 1982). The results obtained from this preliminary experiment of analyzing 34-SNPs included in the Precision ID Identity panel has to observed that Qatar share the same genetic patterns with neighboring regions.

Y chromosome haplogroups determined by the combination of allelic states at binary SNP loci show clearly the different geographic distributions and may predict the ancestral and geographic origins of unknown casework samples (Muro et al., 2011).

The RMP values, relating to the probability of obtaining a match between two distinct and unrelated individuals, provides a useful measure for evaluating the discrimination power of the DNA profiling. In addition, it provides an indication of how rare the genotype is globally (Butler, 2005; Kidd et al., 2018). The plugin provides the RMP values, and the overall RMP range (minimum to maximum). Moreover, the population listed and the top values represent the larger bound for the RMP among the population tested. By reviewing the top geographic regions, RMP results for the 105 Qatari samples were 47.62% (Europe), 39.05% (America), 8.57%, 3.81% (East Asia) and 0.95% (Africa). The 1000 Genome American (indigenous American) comprises three populations namely, Colombian from Medellin, Colombia, Puerto Rican from Puerto Rico and Mexican Ancestry from Los Angeles, California.

On the other hand, the European population is made of Toscani from Italia, Finnish from Finland, British from England and Scotland and Iberian Population from Spain. The South Asian samples are Gujarati Indian from Houston, Texas, Punjabi from Lahore, Pakistan, Bengali from Bangladesh, Sri Lankan Tamil from the UK and Indian Telugu from the UK. East Asian includes Han Chinese from Beijing, China, Japanese in Tokyo, Japan, Chinese Dai in Xishuangbanna, China and Kinh in Ho Chi Minh City, Vietnam and Kinh in Ho Chi Minh City, Vietnam. African is made of Yoruba from Ibadan, Nigeria, Luhya from Webuye, Kenya, Gambian in Western Divisions from Gambia, Mende in Sierra Leone, Esan in Nigeria, Americans of African Ancestry from South West USA and African Caribbean from Barbados.

In the study by Rodriguez-Flores et al., (2016), they reported that when the samples (104 unrelated natives of the Arabian Peninsula who are citizens of the nation of Qatar) were compared to worldwide populations and sampled in the 1000 Genomes Project, the indigenous Arabs had a signal of admixture with Europeans. Moreover, they clustered in a basal, outgroup position to all 1000 Genomes non-Africans. The results place indigenous Arabs as the most distant relatives of all other contemporary non-Africans and identify these people as direct descendants of the first Eurasian populations established by the out-of-Africa migrations (Rodriguez-Flores et al., 2016). Reports for forensic casework analysis RMP were generated from the plugin add values to the case files. However, the 1000 Genomes projects have considered genomes from continents that include Africa, Asia, Europe and North America, but the Arabian genome remains unexplored. Until now, genomic data of populations from Arabian Peninsula are poorly represented in databases (Alsmadi et al., 2014).

A panel of at least 50–100 SNP autosomal loci would be required to obtain the same power of discrimination that existing STR multiplex systems provide (Budowle & Van Daal, 2008). The combined matching probability (CMP), and the combined power of exclusion (CPE) for the 90 autosomal SNPs are included in the Precision ID Identity panel and they were were (7.5674×10^{-37}) and (0.999998714). The CPM for the 90 SNPs is higher than 15 STRs of Identifiler plus 1.90×10^{-20} (Bashir, 2016) and 21 STRs in GlobalFiler 1.42091×10^{-26} (Alsafiah et al., 2019). While it is not the primary intention, the Precision ID Ancestry panel can be used as a tool in forensic DNA testing for identification. The Combined Power of Discrimination (CPD), the Combined Power of Exclusion and the Combined Probability of Match (CPM) were (0.99999), (0.999998) and 1.35×10^{-42} respectively, allowing for a very good level of discrimination in forensic cases. The HID Identity panel is an assay for 90 human SNPs. It could be a helpful tool and give great support to acquire complementary data to improve those obtained with traditional STRs. The combination of both autosomal and Y-SNPs could give more information for the case resolution or be used as investigative tool (García et al., 2017). As the forensic community always is facing in which direction the DNA fingerprint technology will be developed. The availability of MPS platforms and forensic panels that targeting many markers can direct scientists to implement those new promising

technology soon. Also, it can be used in conjunction with gold current STR analysis can help answer unsolved and challenging cases (Haddrill, 2021; Jordan & Mills, 2021; Roewer et al., 2013).

One of the major advantages, for the results obtained in this study from this series of preliminary experiments, showed that the Precision ID Identity and/ or Ancestry panel may be part of forensic casework analysis for human identification. They can also be used to study the derived ancestry parental lineage with the STR standard tool, which may be useful in some cold cases as well as for some samples that may show some issues during the analysis. These include a dropout of the either amelogenin Y allele or in complex cases with mutations, especially when analysis of more than one class of polymorphism is required. This is particular importance when sometimes the laboratory receives both civil and criminal paternity, sometimes with difficult scenarios then analysis using MPS panel may provide valuable information to resolve complex kinship cases.

7.3 Limitations/Challenges of MPS Analysis

Currently, fewer papers have discussed the application of MPS panels on forensic samples or population from regions neighboring Qatar to facilitate any comparison studies between the populations.

For HID SNP Genotyper plugin adding more reference population will improve the analysis process.

7.4 Conclusion

In conclusion, the results have shown that

1. The study provides an excellent start for evaluating the MPS technology in Qatar forensic lab. The availability of MPS platforms and forensic panels allows the possibility to analyse several hundred markers in a single run. MPS results showed a great promise for forensic casework analysis, which might help continue the work on these panels and also to implement these approaches into the routine STR-CE for analysing casework samples. Also, studying the Precision ID Globalfiler™ NGS STR Panel with use of Ion S5 system would be interesting.

2. The allele frequencies and forensic statistical parameters of the Precision ID Identity Panel and Ancestry Panel are reported in Qatari population for the first time.
3. The results collected from the experiments showed that the ability of both panels to be applied for most casework samples, also collection and sequencing of more challenging samples including mixtures and inhibitors would be useful.
4. From the results obtained in this study from both panels, it would be an introduction to start to design threshold through using coverage values, heterozygous balance and noise level.
5. Ion Chef provides was easy to use, it provides an automation system for Automation of the library preparation, template preparation (emulsion PCR and enrichment) and chip loading. The automation of the library building process significantly reduces the hands-on time in the laboratory, decreases the risk of pipetting errors and reduce the workload of the laboratory technicians and also reduce the chance of contamination. It can be suggested that the automated workflow (automated library preparation, automated template preparation and chip loading) is the preferable option when the forensic lab starts implementing MPS.
6. Collecting and sequencing more samples from the population of Qatar using MPS systems and also the results collected would allow establishing a sequence-based database for SNPs beside Qatar national database which is already built on STR.
7. Considering the information collected from both experiments (population and casework samples with Precision ID Ancestry panel) and the addition of Qatari population dataset to the plug-in, it was useful to sort the results. When the population samples were reanalysed with the plug-in included the Qatari samples, some samples were changed and predicted as Qatari and the previous highest likelihood predicted went down, while there were some samples remained unchanged. Moreover, the Qatari population likelihood went to the second or third order and others. From the results obtained in this study, possible suggestions can be made to improve the assignments through some points. They include:-

- a) the results generated from the study open the gate to designing more research on studying more population in Qatar or even the way of sample collection may be in more extended questions and/or including different population resident in the country.
- b) studying and adding reference populations from different regions that are not in the data set could improve the inference process.
- c) in case of admixture present, with the first likelihood rank, the interpretation can be expanded to cover the second and the third order.
- d) some Geo region in HID SNP Genotyper plugin for some populations need to be improve, especially for the populations in Southwest and South Asia.

7.5 Alternative MPS systems

The SNP panels provided by Thermo Fisher have not gained widespread use in the forensic community. A competitor system, ForenSeq, provided by Illumina (now Verogen) that allows simultaneous profiling of STRs and SNPs using MPS has gained more attention. It has the advantage of typing STRs that are the same (or at least overlap) as the STRs analyzed with capillary electrophoresis-based systems. In the context of Qatar, this technology was not available at the time the study was undertaken. Using the Thermo Fisher NGS kit would allow sequencing of the STRs that are currently used on the National DNA database. However, the new information that would be derived would be restricted to sequence variations, which while valuable, do not significantly increase the capacity to discriminate between samples compared to a suite of SNPs designed for this purpose.

The ForenSeq kit by Verogen captures both STRs and SNPs and may be a better option, but is dependent on the availability within of the kits in Qatar.

7.6 Scope for future studies

1. It is necessary to make the system employable in forensic casework more for basic research.

2. Given that the resident population of Qatar is predominantly made up of people from South Asia (Pakistan, India, Nepal and Bangladesh) reference databases relevant to these populations would be valuable, especially if trying to infer ancestry.
3. In addition, an effort should be made to evaluate the efficiency of the panels in order to type real forensic samples. Different types of inhibiting DNA samples with most common inhibitors (humic acid, tannic acid, hematin, ethanol, phenol and EDTA) can be tested to address the issues of DNA inhibition and to give a better understanding about the types of forensic challenged samples which can be successfully typed with the panels to reach for a better evaluation of both panels.
4. Wider population based in Qatar can be included in the future studies and their forensic efficiency and other statistical parameters can be evaluated and that will help to use SNPs whether Identity or Ancestry panel in the routine casework analysis.
5. Through the study, suggestions can be made in studying the possibility of using panels in civil and criminal paternity cases.
6. Bioinformatic analysis by analysing Binary alignment map (BAM) and binary alignment index (BAI) files will be carried out and visualised using the Integrative Genomics Viewer (IGV) to examine flanking sequence variation.

CHAPTER: 8 References

- 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–1073.
- 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56-65.
- 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68-87.
- Abu-Amero, K. K., Hellani, A., González, A. M., Larruga, J. M., Cabrera, V. M., & Underhill, P. A. (2009). Saudi arabian Y-chromosome diversity and its relationship with nearby regions. *BMC Genetics*, *10*, 59-67.
- Al-Asfi, M., McNevin, D., Mehta, B., Power, D., Gahan, M. E., & Daniel, R. (2018). Assessment of the precision ID ancestry panel. *International Journal of Legal Medicine*, *132*, 1581-1594.
- Al-Awadi, S. J., AL-Zubaidi, M. M., Namaa, D. S., Saleh, T. Y., Shehab, M. J., & Hameed, S. N. (2014). Genetic variation of 15 autosomal short tandem repeat (STR) loci in the Diyala–Iraqi population. *International Journal of Biological and Pharmaceutical Research*, *5*(3), 131-135.
- Al-Dosari, W. R., Al-Binali, I. A., Pydi, S. S., & Goodwin, W. (2019). Analysis of forensic casework samples by precision ID ancestry panel – manual and automated ampliseq workflow. *Forensic Science International: Genetics Supplement Series*, *7*(1), 816-817.
- Alsafiah, H. M., Aljanabi, A. A., Hadi, S., Alturayef, S. S., & Goodwin, W. (2019). An evaluation of the SureID 23comp human identification kit for kinship testing. *Scientific Reports*, *9*(1), 1-12.

- Alsmadi, O., John, S. E., Thareja, G., Hebbar, P., Antony, D., Behbehani, K., & Thanaraj, T. A. (2014). Genome at juncture of early human migration: A systematic analysis of two whole genomes and thirteen exomes from Kuwaiti population subgroup of inferred Saudi Arabian tribe ancestry. *PLoS One*, *9*(6), e99069.
- Al-Zahery, N., Pala, M., Battaglia, V., Grugni, V., Hamod, M. A., Kashani, B. H., . . . Semino, O. (2011). In search of the genetic footprints of Sumerians: A survey of Y-chromosome and mtDNA variation in the Marsh Arabs of Iraq. *BMC Evolutionary Biology*, *11*(1), 288.
- Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. (2016). High throughput sequencing: An overview of sequencing chemistry. *Indian Journal of Microbiology*, *56*(4), 394-404.
- Ambers, A., Votrubova, J., Vanek, D., Sajantila, A., & Budowle, B. (2018). Improved Y-STR typing for disaster victim identification, missing persons investigations, and historical human skeletal remains. *International Journal of Legal Medicine*, *132*(6), 1545-1553.
- Amigo, J., Phillips, C., Lareu, M., & Carracedo, Á. (2008). The SNP for ID browser: an online tool for query and display of frequency data from the SNP for ID project. *International Journal of Legal Medicine*, *122*(5), 435-440.
- Araújo, G. S., Lima, L. H. C., Schneider, S., Leal, T. P., da Silva, A. P. C., Vaz de Melo, P. O., ... & Rodrigues, M. R. (2016). Integrating, summarizing and visualizing GWAS-hits and human diversity with DANCE (Disease-ANCEstry networks). *Bioinformatics*, *32*(8), 1247-1249.
- Avila, E., Felkl, A. B., Graebin, P., Nunes, C. P., & Alho, C. S. (2019). Forensic characterization of Brazilian regional populations through massive parallel sequencing of 124 SNPs included in HID ion AmpliSeq identity panel. *Forensic Science International: Genetics*, *40*, 74-84.

- Ballard, D., Winkler-Galicki, J., & Wesolý, J. (2020). Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine*, 134, 1291-1303.
- Bandera, E. V., Chandran, U., Zirpoli, G., Gong, Z., McCann, S. E., Hong, C. C., & Ambrosone, C. B. (2013). Body fatness and breast cancer risk in women of African ancestry. *BMC Cancer*, 13(1), 1-13.
- Bashir, M. (2016). *Application of autosomal INDELS as a forensic tool in Qatar* (Doctoral dissertation, University of Central Lancashire).
- Belmont, J. W., Boudreau, A., Leal, S. M., Hardenbol, P., Pasternak, S., Wheeler, D. A., . . . Stewart, J. (2005). A haplotype map of the human genome. *Nature*, 437(7063), 1299-1320.
- Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., Ch'Ang, L. -, . . . Tanaka, T. (2003). The international HapMap project. *Nature*, 426(6968), 789-796.
- Bhaskar, A., Javanmard, A., Courtade, T. A., & Tse, D. (2017). Novel probabilistic models of spatial genetic ancestry with applications to stratification correction in genome-wide association studies. *Bioinformatics*, 33(6), 879-885.
- Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics*, 18, 78-89.
- Børsting, C. (2015). *Next generation sequencing (NGS) for forensics – methods and definitions. Lecture presented at the Human Identification Ion PGM Forensic Schools, NGS based on STR analysis- training course [Power Point Slide-pdf]*. Darmstadt, Germany.
- Børsting, C., Rockenbauer, E., & Morling, N. (2009a). Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. *Forensic Science International: Genetics*, 4(1), 34-42.

- Børsting, C., Pereira, V., Andersen, J. D., & Morling, N. (2009b). Single nucleotide polymorphism. *Wiley Encyclopedia of Forensic Science*, pp 1-18.
- Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3), 314.
- Brion, M., Sanchez, J. J., Balogh, K., Thacker, C., Blanco-Verea, A., Børsting, C., . . . Schneider, P. M. (2005). Introduction of an single nucleodite polymorphism based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages. *Electrophoresis*, 26(23), 4411-4420.
- Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis*, 39(21), 2642-2654.
- Buchard, A., Kampmann, M., Poulsen, L., Børsting, C., & Morling, N. (2016). ISO 17025 validation of a next-generation sequencing assay for relationship testing. *Electrophoresis*, 37(21), 2822-2831.
- Buckleton JS, Triggs CM, Walsh SJ (2005). Forensic DNA evidence interpretation, 372 CRC Press .
- Budowle, B., & Van Daal, A. (2008). Forensically relevant SNP classes. *BioTechniques*, 44(5), 603-610.
- Budowle, B., Schmedes, S. E., & Wendt, F. R. (2017). Increasing the reach of forensic genetics with massively parallel sequencing. *Forensic Science, Medicine and Pathology*, 13(3), 342-349.
- Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica Et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(10), 1932-1941.

- Butler, J. M., Budowle, B., Gill, P., Kidd, K. K., Phillips, C., Schneider, P. M., . . . Morling, N. (2008). Report on ISFG SNP panel discussion. *Forensic Science International: Genetics Supplement Series*, 1(1), 471-472.
- Butler, J. M. (2005). *Forensic DNA typing: Biology, technology, and genetics of STR markers*. Second Edition ed. Burlington, MA, USA: Elsevier Academic Press.
- Butler, J. M. (2011). *Advanced topics in forensic DNA typing: Methodology*. Academic Press, London, UK.
- Butler, J. M. (2007). Short tandem repeat typing technologies used in human identity testing. *Biotechniques*, 43(4), Sii-Sv.
- Butler, J. M., Coble, M. D., & Vallone, P. M. (2007). STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine, and Pathology*, 3(3), 200-205.
- Cadenas, A. M., Zhivotovsky, L. A., Cavalli-Sforza, L. L., Underhill, P. A., & Herrera, R. J. (2008). Y-chromosome diversity characterizes the gulf of Oman. *European Journal of Human Genetics*, 16(3), 374-386.
- Cann, H. M., De Toma, C., Cazes, L., Legrand, M., Morel, V., Piouffre, L., . . . Cambon-Thomsen, A. (2002). A human genome diversity cell line panel. *Science*, 296(5566), 261-262.
- Canturk, K. M., Emre, R., Kınoglu, K., Başpınar, B., Sahin, F., & Ozen, M. (2014). Current status of the use of single-nucleotide polymorphisms in forensic practices. *Genetic Testing and Molecular Biomarkers*, 18(7), 455-460.
- Carracedo, A. (2005). *Forensic DNA typing protocols*. Totowa, N.J: Humana Press.
Retrieved from <http://capitadiscovery.co.uk/uclan/items/716989>
- Cavalli-Sforza, L. L. (2005). The human genome diversity project: Past, present and future. *Nature Reviews Genetics*, 6(4), 333-340.

- Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., . . . Walsh, S. (2018). *The HirisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation*. *Forensic Science International: Genetics*, 35, 123-135.
- Chial, H. (2008). DNA sequencing technologies key to the human genome project. *Nature Education*, 1(1), 219-225.
- Clair, M. S. (2020). The prehistory of language from the perspective of the Y-chromosome. The Genetic-Linguistic Interface Project. Retrieved from [DOI: 10.13140/RG.2.2.26725.01760](https://doi.org/10.13140/RG.2.2.26725.01760)
- Coble, M. D., Just, R. S., O'Callaghan, J. E., Letmanyi, I. H., Peterson, C. T., Irwin, J. A., & Parsons, T. J. (2004). Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in caucasians. *International Journal of Legal Medicine*, 118(3), 137-146.
- Collins, F. S., Brooks, L. D., & Chakravarti, A. (1998). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research*, 8(12), 1229-1231.
- Corach, D., Risso, L. F., Marino, M., Penacino, G., & Sala, A. (2001). Routine Y-STR typing in forensic casework. *Forensic Science International*, 118(2-3), 131-135.
- Daniel, R., & Walsh, S. J. (2006). The continuing evolution of forensic DNA profiling - from STRs to SNPs. *Australian Journal of Forensic Sciences*, 38(2), 59-74.
- de la Puente, M., Phillips, C., Santos, C., Fondevila, M., Carracedo, Á, & Lareu, M. V. (2017). *Evaluation of the qiagen 140-SNP forensic identification multiplex for massively parallel sequencing*. *Forensic Science International: Genetics*, 28, 35-43.
- Debnath, M., Palanichamy, M. G., Mitra, B., Jin, J., Chaudhuri, T. K., & Zhang, Y. (2011). Y-chromosome haplogroup diversity in the sub-Himalayan terai and Duars populations of east India. *Journal of Human Genetics*, 56(11), 765-771.

- Ding, X., Li, M., Gu, H., Peng, X., Zhang, Z., & Wu, F. (2015). Detecting SNP combinations discriminating human populations from HapMap data. *IEEE Transactions on Nanobioscience*, *14*(2), 220-228.
- Diop, A., Tessler, M., Trung Le, K., Al-Emadi, D., & Howell, D. (2012). Attitudes towards migrant workers in the GCC: Evidence from Qatar. *Journal of Arabian Studies*, *2*(2), 173-187.
- Dixon, L. A., Murray, C. M., Archer, E. J., Dobbins, A. E., Koumi, P., & Gill, P. (2005). *Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes*. *Forensic Science International*, *154*(1), 62-77.
- Eduardoff, M., Gross, T. E., Santos, C., de la Puente, M., Ballard, D., Strobl, C., . . . Phillips, C. (2016). Inter-laboratory evaluation of the EUROFORGEN global ancestry-informative SNP panel by massively parallel sequencing using the ion PGM™. *Forensic Science International: Genetics*, *23*, 178-189.
- Eduardoff, M., Santos, C., de la Puente, M., Gross, T. E., Fondevila, M., Strobl, C., . . . Phillips, C. (2015). Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the ion PGM™. *Forensic Science International: Genetics*, *17*, 110-121.
- Elena, S., Alessandro, A., Ignazio, C., Sharon, W., Luigi, R., & Andrea, B. (2016). *Revealing the challenges of low template DNA analysis with the prototype ion AmpliSeq™ identity panel v2.3 on the PGM™ sequencer*. *Forensic Science International: Genetics*, *22*, 25-36.
- Enoch, M. A., Shen, P. H., Xu, K., Hodgkinson, C., & Goldman, D. (2006). Using ancestry-informative markers to define populations and detect population stratification. *Journal of Psychopharmacology*, *20*(4_suppl), 19-26.
- Espegueira Themudo, G., Smidt Mogensen, H., Børsting, C., & Morling, N. (2016). Frequencies of HID-ion ampliseq ancestry panel markers among greenlanders. *Forensic Science International: Genetics*, *24*, 60-64.

- Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, 10(3), 564-567.
- Fakhro, K. A., Staudt, M. R., Ramstetter, M. D., Robay, A., Malek, J. A., Badii, R., ... & Rodriguez-Flores, J. L. (2016). The Qatar genome: a population-specific tool for precision medicine in the Middle East. *Human Genome Variation*, 3(1), 1-7.
- Fondevila, M., Phillips, C., Santos, C., Freire Aradas, A., Vallone, P. M., Butler, J. M., . . . Carracedo, Á. (2013). Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Science International: Genetics*, 7(1), 63-74.
- Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., . . . Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), 851-861.
- Frudakis, T., Venkateswarlu, K., Thomas, M., Gaskin, Z., Ginjupalli, S., Gunturi, S., . . . Nachimuthu, P. K. (2003). A classifier for the SNP-based inference of ancestry. *Journal of Forensic Sciences*, 48(4), 771-782.
- Fullerton, S. M., & Lee, S. S. (2011). Secondary uses and the governance of de-identified data: Lessons from the human genome diversity panel. *BMC Medical Ethics*, 12(1), 1-6.
- García, O., Soto, A., & Yurrebaso, I. (2017). Allele frequencies and other forensic parameters of the HID-ion AmpliSeq™ identity panel markers in Basques using the ion torrent PGM™ platform. *Forensic Science International: Genetics*, 28, e8-e10.
- García, O., Ajuriagerra, J. A., Alday, A., Alonso, S., Pérez, J. A., Soto, A., . . . Yurrebaso, I. (2017). Frequencies of the precision ID ancestry panel markers in Basques using the ion torrent PGM™ platform. *Forensic Science International: Genetics*, 31, e1-e4.

- Gardner, A. M., Pessoa, S., & Harkness, L. M. (2014). Labour migrants and access to justice in contemporary Qatar. *London School of Economics (LSE), Middle East Center*, pp 1-64.
- Ge, J., Budowle, B., Planz, J. V., & Chakraborty, R. (2010). Haplotype block: A new type of forensic DNA markers. *International Journal of Legal Medicine*, 124(5), 353-361.
- Geserick, G., & Wirth, I. (2012). Genetic kinship investigation from blood groups to DNA markers. *Transfusion Medicine and Hemotherapy*, 39(3), 163-175.
- Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, 18, 9-19.
- Goodwin, W., Linacre, A., & Hadi, S. (2011). *An introduction to forensic genetics*. 2nd ed. Wiley-Blackwell, West Sussex, UK.
- Gopinath, S., Zhong, C., Nguyen, V., Ge, J., Lagacé, R. E., Short, M. L., & Mulero, J. J. (2016). Developmental validation of the YFiler® plus PCR amplification kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic Science International: Genetics*, 24, 164-175.
- Greely, H. T. (2001). Human genome diversity: what about the other human genome project?. *Nature Reviews Genetics*, 2(3), 222-227.
- Grimes, E. A., Noake, P. J., Dixon, L., & Urquhart, A. (2001). Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Science International*, 122(2-3), 124-129.
- Guo, F., Yu, J., Zhang, L., & Li, J. (2017). Massively parallel sequencing of forensic STRs and SNPs using the illumina® ForenSeq™ DNA signature prep kit on the MiSeq FGx™ forensic genomics system. *Forensic Science International: Genetics*, 31, 135-148.

- Guo, F., Zhou, Y., Song, H., Zhao, J., Shen, H., Zhao, B., . . . Jiang, X. (2016). Next generation sequencing of SNPs using the HID-ion AmpliSeq™ identity panel on the ion torrent PGM™ platform. *Forensic Science International: Genetics*, 25, 73-84.d
- Haddrill, P. R. (2021). Developments in forensic DNA analysis. *Emerging Topics in Life Sciences*. doi: <https://doi.org/10.1042/ETLS20200304>
- Halder, I., Shriver, M., Thomas, M., Fernandez, J. R., & Frudakis, T. (2008). A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: Utility and applications. *Human Mutation*, 29(5), 648-658.
- Hammer, M. F., Karafet, T. M., Redd, A. J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., & Zegura, S. L. (2001). Hierarchical patterns of global human Y-chromosome diversity. *Molecular Biology and Evolution*, 18(7), 1189-1203.
- Hammer, M. F., Karafet, T., Rasanayagam, A., Wood, E. T., Altheide, T. K., Jenkins, T., ... & Zegura, S. L. (1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Molecular Biology and Evolution*, 15(4), 427-441.
- Hammer, M. F., & Zegura, S. L. (2002). The human Y chromosome haplogroup tree: Nomenclature and phylogeography of its major divisions. *Annual Review of Anthropology*, 31(1), 303-321.
- Hammer, M., & Redd, A. J. (2006). Forensic applications of Y chromosome STRs and SNPs. *Forensics in Law Enforcement*, pp 133.
- Hanson, E. K., & Ballantyne, J. (2007). An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system. *PloS one*, 2(8), e688.
- Hares, D. R. (2015). Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International. Genetics*, 17, 33-34.
- Harrington, C. T., Lin, E. I., Olson, M. T., & Eshleman, J. R. (2013). Fundamentals of pyrosequencing. *Arch Pathology Laboratory Medicine*, 137(9), 1296-1303.

- Hollard, C., Keyser, C., Delabarde, T., Gonzalez, A., Lamego, C. V., Zvéni gorosky, V., & Ludes, B. (2017). Case report: On the use of the HID-ion AmpliSeq™ ancestry panel in a real forensic case. *International Journal of Legal Medicine*, 131(2), 351-358.
- Hood, L., & Rowen, L. (2013). The human genome project: big science transforms biology and medicine. *Genome Medicine*, 5(9), 1-8.
- Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931-945.
- Huston, K. A. (1998). Statistical analysis of STR data. *Profiles DNA*, 1(3), 14-15.
- Illumina. (2016). *ForenSeq™ DNA signature prep Kit Guide*. vol 15049528, v01 edn, CA, USA.
- Illumina. (2015). *ForenSeq™ DNA Signature Prep Guide, P/N 15049528, Rev. C*, Illumina, San Diego.
- International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311), 52.
- Iozzi, S., Carboni, I., Contini, E., Pescucci, C., Frusconi, S., Nutini, A. L., . . . Ricci, U. (2015). *Forensic genetics in NGS era: New frontiers for massively parallel typing*. *Forensic Science International: Genetics Supplement Series*, 5, e418-e419.
- Ip, S. C., Lin, S. W., Li, C., & Lai, K. M. (2014). Forensic DNA typing strategy of degraded DNA on discarded cigarette ends using the AmpFℓSTR® Identifiler®, Identifiler® Plus and MiniFiler™ PCR Amplification Kits. *Science & Justice*, 54(4), 311-315.
- Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., . . . Caves, G. (2017). Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Science International: Genetics*, 28, 52-70.

- Jeffrey, M. P., & Peter, A. F. (2016). Next-gen sequencing 2016 update. Retrieved from <https://www.biocompare.com/Editorial-Articles/183239-Next-Gen-Sequencing-2016-Update/>
- Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985a). Hypervariable 'minisatellite' regions in human DNA. *Nature*, *314*(6006), 67-73.
- Jeffreys, A. J., Brookfield, J. F., & Semeonoff, R. (1985b). Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, *317*(6040), 818-819.
- Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, *5*(10), 739-751.
- Jobling, M. A. (2001). Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Science International*, *118*(2-3), 158-162.
- Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker comes of age. *Nature Reviews Genetics*, *4*(8), 598-612.
- Jobling, M., Hurles, M., & Tyler-Smith, C. (2013). *Human evolutionary genetics: Origins, peoples & disease*. Garland Science. New York.
- Jordan, D., & Mills, D. (2021). Past, present, and future of DNA typing for analyzing human and non-human forensic samples. *Frontiers in Ecology and Evolution*, *9*, 172.
- Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research*, *18*(5), 830-838.
- Kaushik, A., Sharma, A., & Sharma, S. (2018). Haplotype: Alleles moving together. *Journal of Pharmacognosy and Phytochemistry*, *7*(2), 2334-2236.
- Kayser, M. (2017). Forensic use of Y-chromosome DNA: a general overview. *Human Genetics*, *136*(5), 621-635.

- Kchouk, M., Gibrat, J., & Elloumi, M. (2017). Generations of sequencing technologies: From first to next generation. *Biology and Medicine*, *9*(3), 1-8.
- Kersbergen, P., van Duijn, K., Kloosterman, A. D., den Dunnen, J. T., Kayser, M., & de Knijff, P. (2009). Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC Genetics*, *10*:69.
- Kidd, K. K., Pakstis, A. J., Speed, W. C., Grigorenko, E. L., Kajuna, S. L., Karoma, N. J., ... & Kidd, J. R. (2006). Developing a SNP panel for forensic identification of individuals. *Forensic Science International*, *164*(1), 20-32.
- Kidd, K. K., Speed, W. C., Pakstis, A. J., Furtado, M. R., Fang, R., Madbouly, A., . . . Kidd, J. R. (2014). Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Science International: Genetics*, *10*, 23-32.
- Kidd, K. K., Kidd, J. R., Speed, W. C., Fang, R., Furtado, M. R., Hyland, F., & Pakstis, A. J. (2012). Expanding data and resources for forensic use of SNPs in individual identification. *Forensic Science International: Genetics*, *6*(5), 646-652.
- Kidd, K. K., Soundararajan, U., Rajeevan, H., Pakstis, A. J., Moore, K. N., & Roper-Miller, J. D. (2018). The redesigned forensic research/reference on genetics-knowledge base, FROG-kb. *Forensic Science International: Genetics*, *33*, 33-37.
- Kim, S., & Misra, A. (2007). SNP genotyping: technologies and biomedical applications. *Annual Rev. Biomedical Engineering*, *9*, 289-320.
- Klyosov, A. A., & Rozhanskii, I. L. (2012). Haplogroup R1a as the proto indo-europeans and the legendary Aryans as witnessed by the DNA of their current descendants. *Advances in Anthropology*, *2*(01), 1-13.
- Köcher, S., MÃ¼ller, P., Berger, B., Bodner, M., Parson, W., Roewer, L., . . . DNASEqEx Consortium. (2018). Inter-laboratory validation study of the ForenSeq DNA signature prep kit. *Forensic Science International: Genetics*, *36*, 77-85.
- Kosoy, R., Nassir, R., Tian, C., White, P. A., Butler, L. M., Silva, G., . . . Belmont, J. W. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation*, *30*(1), 69-78.

- Kulski, J. K. (2016). Next-generation sequencing—an overview of the history, tools, and “Omic” applications. *Next Generation Sequencing—Advances, Applications and Challenges, Intech* , pp 3-60.
- Kwok, P. (2001). Methods for genotyping single nucleotide polymorphisms. *Annual Review Genomics Human Genetics, 2*, 235-258.
- Lamason, R. L., Mohideen, M. P., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., . . . Humbert, J. E. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science, 310*(5755), 1782-1786.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . FitzHugh, W. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*, 860-921.
- Landsteiner, K. (1900). Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralbl. Bakteriol., 27*, 357-362.
- Lao, O., Vallone, P. M., Coble, M. D., Diegoli, T. M., Van Oven, M., Van Der Gaag, Kristiaan J, . . . Kayser, M. (2010). Evaluating self-declared ancestry of US americans with autosomal, y-chromosomal and mitochondrial DNA. *Human Mutation, 31*(12), e1875-e1893.
- Lao, O., van Duijn, K., Kersbergen, P., de Knijff, P., & Kayser, M. (2006). Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *The American Journal of Human Genetics, 78*(4), 680-690.
- LaRue, B. L., Ge, J., King, J. L., & Budowle, B. (2012). A validation study of the qiagen investigator DIPplex[®] kit; an INDEL-based assay for human identification. *International Journal of Legal Medicine, 126*(4), 533-540.
- Laurent, F. X. (2018). Implementation of a fully automated workflow using MPS technology in a high volume casework laboratory. *Paper presented at the DNA Evidence to Investigative Insights DEII 2018*. Barcelona, Spain.

- Lee, J. H., Cho, S., Kim, M., Shin, D. H., Rakha, A., Shinde, V., & Lee, S. D. (2018). Genetic resolution of applied biosystems™ precision ID ancestry panel for seven asian populations. *Legal Medicine*, *34*, 41-47.
- Lessig, R., Zoledziwska, M., Fahr, K., Edelmann, J., Kostrzewa, M., Dobosz, T., & Kleemann, W. J. (2005). Y-SNP-genotyping—a new approach in forensic analysis. *Forensic Science International*, *154*(2-3), 128-136.
- Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., . . . Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, *319*(5866), 1100-1104.
- Li, R., Zhang, C., Li, H., Wu, R., Li, H., Tang, Z., . . . Wang, Y. (2018). SNP typing using the HID-ion AmpliSeq™ identity panel in a southern Chinese population. *International Journal of Legal Medicine*, *132*(4), 997-1006.
- Life Technologies Corporation. 2011. Ion Torrent Amplicon Sequencing (Application Note CO22802). Retrieved from http://application.sb-roscoff.fr/download/fr2424/abims/corre/ngs/3_life/amplicon_application_note_040411.pdf
- Liu, J., Wang, Z., He, G., Zhao, X., Wang, M., Luo, T., . . . Hou, Y. (2018). Massively parallel sequencing of 124 SNPs included in the precision ID identity panel in three east asian minority ethnicities. *Forensic Science International: Genetics*, *35*, 141-148.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., . . . Law, M. (2012). Comparison of next-generation sequencing systems. *BioMed Research International*, *2012*, article 251364
- Ludeman, M. J., Zhong, C., Mulero, J. J., Lagacé, R. E., Hennessy, L. K., Short, M. L., & Wang, D. Y. (2018). Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. *International Journal of Legal Medicine*, *132*(6), 1555-1573.
- Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C., . . . Herrera, R. J. (2004). The levant versus the horn of africa: Evidence for

- bidirectional corridors of human migrations. *The American Journal of Human Genetics*, 74(3), 532-544.
- Liu, Y. Y., & Harbison, S. (2018). A review of bioinformatic methods for forensic DNA analyses. *Forensic Science International: Genetics*, 33, 117-128.
- Malyarchuk, B., Derenko, M., Wozniak, M., & Grzybowski, T. (2013). Y-chromosome variation in Tajiks and Iranians. *Annals of Human Biology*, 40(1), 48-54.
- Marano, L. A., & Fridman, C. (2019). DNA phenotyping: Current application in forensic science. *Research and Reports in Forensic Medical Science*, 9, 1-8.
- Martín, P., de Simón, L. F., Luque, G., Farfán, M. J., & Alonso, A. (2014). Improving DNA data exchange: Validation studies on a single 6 dye STR kit with 24 loci. *Forensic Science International: Genetics*, 13, 68-78.
- McCord, B. R., Gauthier, Q., Cho, S., Roig, M. N., Gibson-Daw, G. C., Young, B., . . . Lee, S. B. (2018). Forensic DNA analysis. *Analytical Chemistry*, 91(1), 673-688.
- Meiklejohn, K. A., & Robertson, J. M. (2017). Evaluation of the precision ID identity panel for the ion torrent™ PGM™ sequencer. *Forensic Science International: Genetics*, 31, 48-56.
- Ministry of Development Planning and Statistics. (2015). *Qatar in figures MDPS AE 2015*. (30). Doha-Qatar: pp 1-78
- Ministry of Development Planning and Statistics. (2016a). *Analysis of the results of population, housing and establishments census 2015*. Doha – Qatar: 1-68.
- Ministry of Development Planning and Statistics. (2016b). *First section population and social statistics*. The State of Qatar, Doha: pp 1-20.
- Ministry of Development Planning and Statistics. (2017). *Births & Deaths in the State of Qatar, 2016 (Review & Analysis)*. The State of Qatar, Doha: pp 1-45.
- Ministry of Foreign Affairs. (2019). Retrieved from <https://www.mofa.gov.qa/en/qatar/history-of-qatar/historical-overview>
- Ministry of Interior of Qatar. (2018). Entry permits (visas). welcome to qatar. Retrieved from <https://portal.moi.gov.qa/qatarvisas/>

- Mizuno, N., Kitayama, T., Fujii, K., Nakahara, H., Yoshida, K., Sekiguchi, K., . . . Kasai, K. (2010). A forensic method for the simultaneous analysis of biallelic markers identifying Y chromosome haplogroups inferred as having originated in asia and the japanese archipelago. *Forensic Science International: Genetics*, 4(2), 73-79.
- Mogensen, H. S., Børsting, C., & Morling, N. (2015). Template preparation of AmpliSeq™ libraries using the ion chef™. *Forensic Science International: Genetics Supplement Series*, 5, e368-e369.
- Mountain, J. L., Knight, A., Jobin, M., Gignoux, C., Miller, A., Lin, A. A., & Underhill, P. A. (2002). SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Research*, 12(11), 1766-1772.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R. K., Horn, G. T., & Erlich, H. (1986). (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Paper presented at the *Cold Spring Harbor Symposia on Quantitative Biology*, , 51 263-273.
- Muro, T., Iida, R., Fujihara, J., Yasuda, T., Watanabe, Y., Imamura, S., . . . Toga, T. (2011). Simultaneous determination of seven informative Y chromosome SNPs to differentiate east Asian, European, and African populations. *Legal Medicine*, 13(3), 134-141.
- Nakano, M., Komatsu, J., Matsuura, S. I., Takashima, K., Katsura, S., & Mizuno, A. (2003). Single-molecule PCR using water-in-oil emulsion. *Journal of biotechnology*, 102(2), 117-124.
- Nievergelt, C. M., Maihofer, A. X., Shekhtman, T., Libiger, O., Wang, X., Kidd, K. K., & Kidd, J. R. (2013). Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. *Investigative Genetics*, 4(1), 1-16.
- Obaidan, Y. M. (1982). System of government in gulf countries. A comparative study of Qatar, Kuwait and Bahrain. *Dar Al-Kutub- Qatar*. (In Arabic).

- Ochiai, E., Minaguchi, K., Nambiar, P., Kakimoto, Y., Satoh, F., Nakatome, M., . . . Osawa, M. (2016). *Evaluation of Y chromosomal SNP haplogrouping in the HID-ion AmpliSeq™ identity panel. Legal Medicine, 22*, 58-61.
- Omberg, L., Salit, J., Hackett, N., Fuller, J., Matthew, R., Chouchane, L., . . . Mezey, J. G. (2012). Inferring genome-wide patterns of admixture in Qataris using fifty-five ancestral populations. *BMC Genetics, 13*: 49.
- Oostdik, K., Lenz, K., Nye, J., Schelling, K., Yet, D., Bruski, S., . . . Storts, D. R. (2014). *Developmental validation of the PowerPlex® fusion system for analysis of casework and reference samples: A 24-locus multiplex for new database standards. Forensic Science International: Genetics, 12*, 69-76.
- Pakstis, A. J., Speed, W. C., Kidd, J. R., & Kidd, K. K. (2007). Candidate SNPs for a universal individual identification panel. *Human Genetics, 121*(3-4), 305-317.
- Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics, 127*(3), 315-324.
- Pardo-Seco, J., Martínón-Torres, F., & Salas, A. (2014). Evaluating the accuracy of AIM panels at quantifying genome ancestry. *BMC Genomics, 15*: 543-548
- Parson, W. (2018). Age estimation with DNA: From forensic DNA fingerprinting to forensic (epi) genomics: A mini-review. *Gerontology, 64*(4), 326-332.
- Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., & Phillips, C. (2016). Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics, 22*, 54-63.
- Pereira, V., Mogensen, H. S., Børsting, C., & Morling, N. (2017). Evaluation of the precision ID ancestry panel for crime case work: A SNP typing assay developed for typing of 165 ancestral informative markers. *Forensic Science International: Genetics, 28*, 138-145.

- Pérez-Miranda, A. M., Alfonso-Sánchez, M. A., Peña, J. A., & Herrera, R. J. (2006). Qatari DNA variation at a crossroad of human migrations. *Human Heredity*, 61(2), 67-79.
- Phillips, C. (2012). Applications of autosomal SNPs and indels in forensic analysis. *Forensic Science Review*, 24(1), 44-62.
- Phillips, C., Amigo, J., Carracedo, Á, & Lareu, M. V. (2015). *Tetra-allelic SNPs: Informative forensic markers compiled from public whole-genome sequence data. Forensic Science International: Genetics*, 19, 100-106.
- Phillips, C., Amigo, J., Tillmar, A. O., Peck, M. A., de la Puente, M., Ruiz-Ramírez, J., ... & Lareu, M. V. (2020). A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel. *Forensic Science International: Genetics*, 46, 102232.
- Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, A., Álvarez-Dios, J., . . . Carracedo, Á. (2007a). Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics*, 1(3-4), 273-280.
- Phillips, C., Fang, R., Ballard, D., Fondevila, M., Harrison, C., Hyland, F., . . . Sobrino, B. (2007b). Evaluation of the genplex SNP typing system and a 49plex forensic marker panel. *Forensic Science International: Genetics*, 1(2), 180-185.
- Phillips, C., Fondevila, M., & Lareau, M. V. (2012). A 34-plex autosomal SNP single base extension assay for ancestry investigations. *DNA electrophoresis protocols for forensic genetics* (pp. 109-126).
- Phillips, C., Lareu, V., Salas, A., & Carracedo, A. (2004). *Nonbinary single-nucleotide polymorphism markers. International Congress Series* ,1261,27-29.
- Phillips, C., Prieto, L., Fondevila, M., Salas, A., Gómez-Tato, A., Álvarez-Dios, J., . . . Montesino, M. (2009). Ancestry analysis in the 11-M madrid bomb attack investigation. *PloS One*, 4(8), e6583.

- Phillips, C., Aradas, A. F., Kriegel, A. K., Fondevila, M., Bulbul, O., Santos, C., ... & Lareu, M. V. (2013). Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries. *Forensic Science International: Genetics*, 7(3), 359-366.
- Phillips, C., McNevin, D., Kidd, K. K., Lagacé, R., Wootton, S., De La Puente, M., ... & Daniel, R. (2019). MAPlex-A massively parallel sequencing ancestry analysis multiplex for Asia-Pacific populations. *Forensic Science International: Genetics*, 42, 213-226.
- Phipps, M. (2015). *PGM for HID: Technology overview* [PowerPoint slides]. Thermo Fisher Scientific. Pp 1-6
- Planning and Statistics Authority. (2019a). *Labor force sample survey: Annual Report 2018*. Doha-Qatar. Retrieved from https://www.psa.gov.qa/en/statistics/Statistical%20Releases/Social/LaborForce/2018/statistical_analysis_labor_force_2018_En.pdf
- Planning and Statistics Authority. (2019b). *Qatar second national development strategy 2018-2022*. Doha-Qatar. Retrieved from <https://www.psa.gov.qa/en/knowledge/Documents/NDS2Final.pdf>
- Pritchard, J. K., Wen, W., & Falush, D. (2010). *Documentation for STRUCTURE software: Version 2.3*. Chicago: University of Chicago, 39.
- Quintana-Murci, L., Krausz, C., & McElreavey, K. (2001). The human Y chromosome: Function, evolution and disease. *Forensic Science International*, 118(2-3), 169-181.
- Rajeevan, H., Soundararajan, U., Pakstis, A. J., & Kidd, K. K. (2012). Introducing the forensic research/reference on genetics knowledge base, FROG-kb. *Investigative Genetics*, 3(1), 18-25.
- Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-throughput sequencing technologies. *Molecular Cell*, 58(4), 586-597.
- Riman, S., Iyer, H., Borsuk, L. A., & Vallone, P. M. (2020). *Understanding the characteristics of sequence-based single-source DNA profiles*. *Forensic Science International: Genetics*, 44, Article 102192.

- Rodriguez-Flores, J. L., Fakhro, K., Agosto-Perez, F., Ramstetter, M. D., Arbiza, L., Vincent, T. L., . . . Chouchane, L. (2016). Indigenous arabs are descendants of the earliest split from ancient Eurasian populations. *Genome Research*, 26(2), 151-162.
- Roewer, L. (2013). DNA fingerprinting in forensics: past, present, future. *Investigative Genetics*, 4(1), 1-10.
- Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, 298(5602), 2381-2385.
- Rosenberg, N. A., Li, L. M., Ward, R., & Pritchard, J. K. (2003). Informativeness of genetic markers for inference of ancestry. *The American Journal of Human Genetics*, 73(6), 1402-1422.
- Rozhanskii, I. L., & Klyosov, A. A. (2012). Haplogroup R1a, its subclades and branches in Europe during the last 9,000 years. *Advances in Anthropology*, 2(03), 139-144.
- Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., . . . Syndercombe-Court, D. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*, 27(9), 1713-1724.
- Santangelo, R., González-Andrade, F., Børsting, C., Torroni, A., Pereira, V., & Morling, N. (2017). *Analysis of ancestry informative markers in three main ethnic groups from ecuador supports a trihybrid origin of ecuadorians. Forensic Science International: Genetics*, 31, 29-33. doi:<https://doi.org/10.1016/j.fsigen.2017.08.012>
- Schneider, P. M. (2009). Expansion of the European standard set of DNA database loci—the current situation. *Profiles in DNA*, 12, 6-7.
- Serrote, C. M. L., Reiniger, L. R. S., Silva, K. B., Rabaiolli, Silvia Machado dos Santos, & Stefanel, C. M. (2020). Determining the polymorphism information content of a molecular marker. *Gene*, 726, 144175.

- Seton, C. (1988). Life for sex killer who sent decoy to take genetic test. *The Times*, January 23, 1988. London.
- Shamsuddin, S. M., & Ahmad, S. S. B. H. (2020). Ancient History of Arabian Peninsula and Semitic Arab Tribes. *Advances in Social Sciences Research Journal*, 7(5), 270-282.
- Shrivastava, P., Jain, T., & Trivedi, V. B. (2016). DNA fingerprinting: A substantial and imperative aid to forensic investigation. *European Journal of Forensic Science*, 3(3), 23-30.
- Shriver, M. D., Parra, E. J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Kittles, R. A. (2003). Skin pigmentation, biogeographical ancestry and admixture mapping. *Human Genetics*, 112(4), 387-399.
- Smigielski, E. M., Sirotkin, K., Ward, M., & Sherry, S. T. (2000). dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Research*, 28(1), 352-355.
- Sobrinho, B., Brión, M., & Carracedo, A. (2005). SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Science International*, 154(2-3), 181-194.
- Syvänen, A. (2001). Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2(12), 930-942.
- Tamaki, K., & Jeffreys, A. J. (2005). Human tandem repeat sequences in forensic DNA typing. *Legal Medicine*, 7(4), 244-250.
- Tan, Y., Bai, P., Wang, L., Wang, H., Tian, H., Jian, H., . . . Zhang, L. (2018). Two-person DNA mixture interpretation based on a novel set of SNP-STR markers. *Forensic Science International: Genetics*, 37, 37-45.
- Tereba, A. (1999). Tools for analysis of population statistics. *Profiles DNA*, 2, 14-16.
- The Permanent Population Committee. (2009). *The population of the state of Qatar 2009*. The State of Qatar, Doha.: pp 1-49.

Thermo Fisher Scientific (2010). *PrepFiler Express™ and PrepFiler Express BTA™ Forensic DNA Extraction Kits USER GUIDE*. Part Number 4442699 Rev. B.
http://tools.thermofisher.com/content/sfs/brochures/cms_086049.pdf

Thermo Fisher Scientific. (2015). Ion PGM system specification sheet.
<https://tools.thermofisher.com/content/sfs/brochures/PGM-Specification-Sheet.pdf>

Thermo Fisher Scientific. (2015). *Ion S5 and ion S5 XL systems specification sheet*.
<https://www.thermofisher.com/content/dam/LifeTech/Documents/PDFs/Ion-S5XL-brochure.pdf>

Thermo Fisher Scientific. (2015). *Ion AmpliSeq™ library preparation for human identification. USER GUIDE, Publication Number MAN0010640 Revision C.0, 1-46*.
https://tools.thermofisher.com/content/sfs/manuals/MAN0010640_AmpliSeqLibraryPrep_HID_UG.pdf

Thermo Fisher Scientific (2016). *Ion Chef System for human identification*.
<https://assets.thermofisher.com/TFS-Assets/LSG/brochures/ion-chef-system-brochure.pdf>

Thermo Fisher Scientific. (2016a). *Precision ID ancestry panel flyer*.
<https://www.thermofisher.com/content/dam/LifeTech/Documents/PDFs/HID-Precision-ID-Ancestry-Panel-Flyer.pdf>

Thermo Fisher Scientific. (2016b). *Precision ID identity panel flyer*.
<https://www.thermofisher.com/content/dam/LifeTech/Documents/PDFs/HID-Precision-ID-Identity-Panel-Flyer.pdf>

Thermo Fisher Scientific. (2017a). *Expand your forensics workflow with the precision ID NGS system for human identification*.
<https://www.thermofisher.com/content/dam/LifeTech/Documents/PDFs/HID-Precision-ID-System-Brochure.pdf>

Thermo Fisher Scientific. (2017b). *HID SNP genotyper plugin- USER GUIDE. v5.2.2, publication number MAN0010641 Revision D.0*.

https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0010641_HIDSNP_Genotyper_Plugin.pdf

Thermo Fisher Scientific (2017). *Precision ID panels with the Ion PGM™ System Application Guide*. Revision B.0. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0015830_PrecisionID_Panels_IonPGM_UG.pdf

Thermo Fisher Scientific (2018). *Quantifiler, HP Trio DNA Quantification Kits User Guide*, Publication No. 4485354 Rev. H <https://assets.thermofisher.com/TFS-Assets/LSG/manuals/4485354.pdf>

Theyab, J. (2013). The genetic structure of the Kuwaiti and Failaka island populations: Y-chromosome & mitochondrial DNA variation (Doctoral dissertation, University of Kansas).

Thorisson, G. A., & Stein, L. D. (2003). The SNP consortium website: Past, present and future. *Nucleic Acids Research*, *31*(1), 124-127.

Tobler, A. R., Short, S., Andersen, M. R., Paner, T. M., Briggs, J. C., Lambert, S. M., . . . Wenz, H. M. (2005). The SNPlex genotyping system: A flexible and scalable platform for SNP genotyping. *Journal of Biomolecular Techniques*, *16*(4), 398-406.

Toma, T. T., Dawson, J. M., & Adjeroh, D. A. (2018). Human ancestry identification under resource constraints--what can one chromosome tell us about human biogeographical ancestry?. *BMC Medical Genomics*, *11*(5), 75-87.

Tomas, C., Stangegaard, M., Børsting, C., Hansen, A. J., Morling, N., & SNPforID Consortium. (2008). Typing of 48 autosomal SNPs and amelogenin with GenPlex SNP genotyping system in forensic genetics. *Forensic Science International: Genetics*, *3*(1), 1-6.

Triki-Fendri, S., Sánchez-Diz, P., Rey-González, D., Alfadhli, S., Ayadi, I., Ben Marzoug, R., . . . Rebai, A. (2016). Genetic structure of the Kuwaiti population revealed by paternal lineages. *American Journal of Human Biology*, *28*(2), 203-212.

Turchi, C., Onofri, V., Melchionda, F., Bini, C., Previderè, C., Carnevali, E., . Tagliabracci, A. (2019). *Dealing with low amounts of degraded DNA: Evaluation of SNP typing of*

challenging forensic samples by using massive parallel sequencing. Forensic Science International: Genetics Supplement Series, 7(1), 83-84.

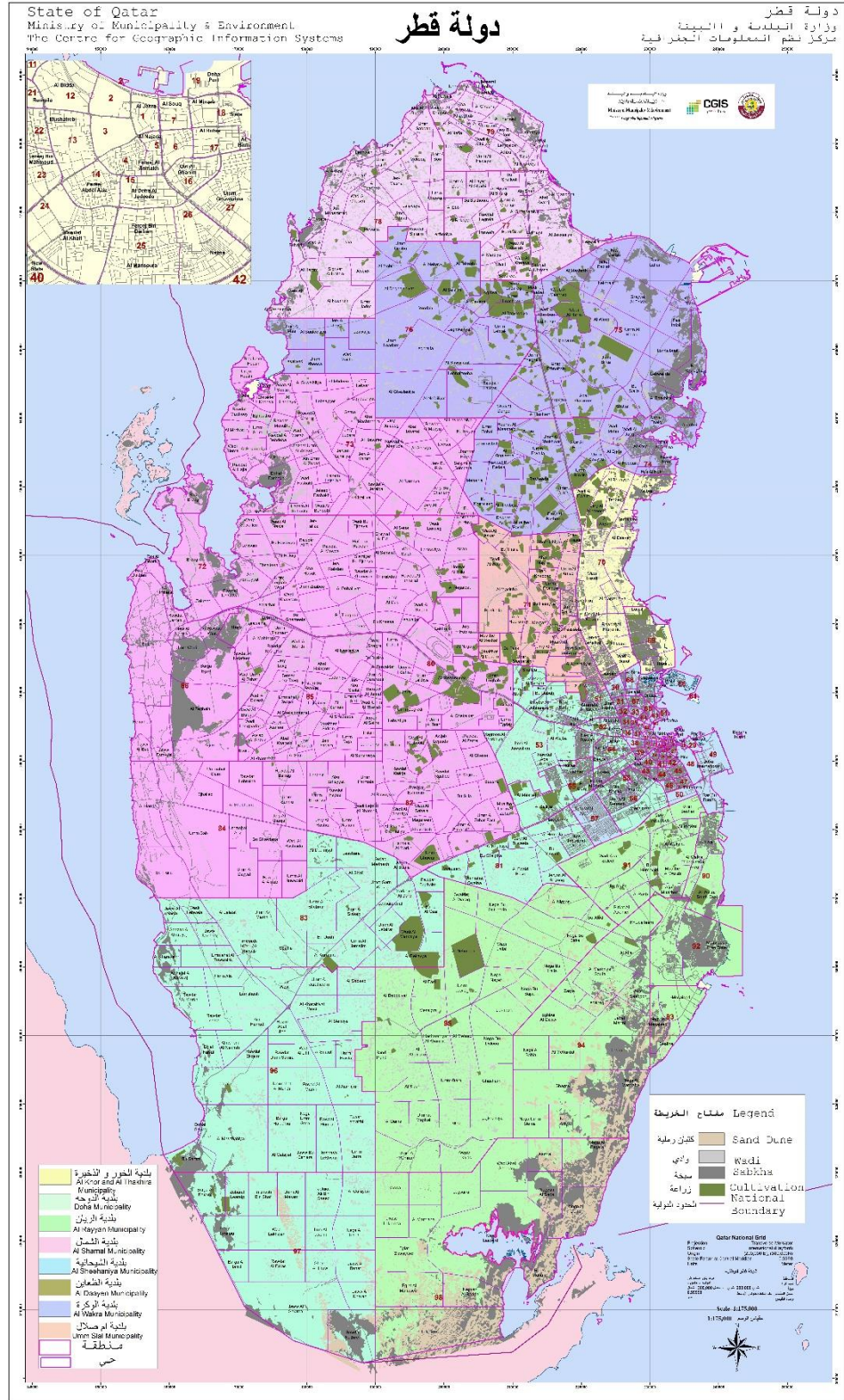
- Twyman, R. M. (2009). Single-nucleotide polymorphism (SNP) analysis. *Encyclopedia of Neurosci, 8, 881-885.*
- Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Lahr, M. M., Foley, R. A., ... & Cavalli-Sforza, L. L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of human genetics, 65(1), 43-62.*
- Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., ... & Oefner, P. J. (2000). Y chromosome sequence variation and the history of human populations. *Nature Genetics, 26(3), 358-361.*
- van der Heijden, S., de Oliveira, S. J., Kampmann, M., Børsting, C., & Morling, N. (2017). *Comparison of manual and automated AmpliSeq™ workflows in the typing of a Somali population with the precision ID identity panel. Forensic Science International: Genetics, 31, 118-125.*
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics, 30(9), 418-426.*
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., . . . Zhu, X. (2001). The sequence of the human genome. *Science, 291(5507), 1304-1351.*
- Vierstraete, A. (2012). Next generation sequencing for dummies. *University Lecture, June 2012.*
- Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., . . . Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics, 7(1), 98-115.*
- Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., de Knijff, P., Kayser, M., & Ballantyne, K. N. (2011a). Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics, 5(5), 464-471.*

- Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O., & Kayser, M. (2011b). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics*, 5(3), 170-180.
- Wang, L., Chen, M., Wu, B., Liu, Y., Zhang, G., Jiang, L., . . . Ye, J. (2018). Massively parallel sequencing of forensic STR s using the ion chef™ and the ion S5™ XL systems. *Journal of Forensic Sciences*, 63(6), 1692-1703.
- Wang, Z., He, G., Luo, T., Zhao, X., Liu, J., Wang, M., . . . Hou, Y. (2018). Massively parallel sequencing of 165 ancestry informative SNPs in two Chinese Tibetan-Burmese minority ethnicities. *Forensic Science International: Genetics*, 34, 141-147.
- Waterston, R. H., Lander, E. S., & Sulston, J. E. (2002). On the sequencing of the human genome. *Proceedings of the National Academy of Sciences*, 99(6), 3712-3716.
- Wen, W., Shu, X. O., Guo, X., Cai, Q., Long, J., Bolla, M. K., ... & Zheng, W. (2016). Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry. *Breast Cancer Research*, 18(1), 1-8.
- Westen, A. A., Matai, A. S., Laros, J. F., Meiland, H. C., Jasper, M., de Leeuw, W. J., ... & Sijen, T. (2009). Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. *Forensic Science International: Genetics*, 3(4), 233-241.
- Wetton, J. H., Tsang, K. W., & Khan, H. (2005). Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs. *Forensic Science International*, 152(1), 45-53.
- Wyner, N., Barash, M., & McNevin, D. (2020). Forensic Autosomal Short Tandem Repeats and Their Potential Association With Phenotype. *Frontiers in Genetics*, 11, 884.
- Y Chromosome Consortium. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research*, 12(2), 339-348.

- Yang, Y., Xie, B., & Yan, J. (2014). Application of next-generation sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics*, 12(5), 190-197.
- Zaykin, D. V., Pudovkin, A., & Weir, B. S. (2008). Correlation-based inference for linkage disequilibrium with multiple alleles. *Genetics*, 180(1), 533-545.
- Zgonjanin, D., Soler, M. P., Antov, M., Redhead, P., Stojiljković, G., & Milić, A. (2017). Validation and implementation of the Investigator® 24plex QS kit for forensic casework. *Forensic Science International: Genetics Supplement Series*, 6, e77-e79.
- Zhang, S., Wang, Z., Bian, Y., Li, L., Hou, Y., & Li, C. (2015). Evaluation of HID-ion Ampliseq™ panel in HAN population. *Forensic Science International: Genetics Supplement Series*, 5, e584-e586.
- Zheng-Bradley, X., & Flicek, P. (2017). Applications of the 1000 genomes project resources. *Briefings in Functional Genomics*, 16(3), 163-170.
- Ziętkiewicz, E., Witt, M., Daca, P., Żebracka-Gala, J., Goniewicz, M., Jarząb, B., & Witt, M. (2012). Current genetic methodologies in the identification of disaster victims and in forensic analysis. *Journal of Applied Genetics*, 53(1), 41-60.

CHAPTER: 9 APPENDICES

9.1 Appendix 1: Map of Qatar



9.2 Appendix 2: Ethical approval letter from the University of Central Lancashire's STEM Ethics Committee.



15th February 2016

William H Goodwin/Waad Rashed Sh a Al-dossari
School of Forensic and Investigative Sciences
University of Central Lancashire

Dear William/Waad,

Re: STEMH Ethics Committee Application
Unique Reference Number: STEMH 371

The STEMH ethics committee has granted approval of your proposal application 'Evaluation of the HID-Ion Ampliseq Ancestry Panel in the Qatar population and forensic casework samples by next generation sequencing using the Ion Torrent PGM'. Approval is granted up to the end of project date* or for 5 years from the date of this letter, whichever is the longer.

It is your responsibility to ensure that:

- the project is carried out in line with the information provided in the forms you have submitted
- you regularly re-consider the ethical issues that may be raised in generating and analysing your data
- any proposed amendments/changes to the project are raised with, and approved, by Committee
- you notify roffice@uclan.ac.uk if the end date changes or the project does not start
- serious adverse events that occur from the project are reported to Committee
- a closure report is submitted to complete the ethics governance procedures (Existing paperwork can be used for this purposes e.g. funder's end of grant report; abstract for student award or NRES final report. If none of these are available use [e-Ethics Closure Report Proforma](#)).

Yours sincerely,

A handwritten signature in black ink, appearing to read "Colin Thain", is written over a light grey rectangular background.

Colin Thain
Chair
STEMH Ethics Committee

* for research degree students this will be the final lapse date

NB - Ethical approval is contingent on any health and safety checklists having been completed, and necessary approvals as a result of gained.

9.3 Appendix 3: Explanation of the Hardy –Weinberg equilibrium, pairwise F_{ST} , linkage disequilibrium (LD), and forensic parameters.

1. The Hardy-Weinberg equilibrium (HWE) predicts the stability of allele and genotype frequencies from one generation to another at one genetic locus. When a population is in HWE, the genotype frequencies can be predicted from the allele frequencies. The genotype frequency of a homozygous can be calculated using p^2 and $2pq$ is used in case of heterozygous genotype frequency calculation (Goodwin et. al. 2011)
2. The F_{ST} is one of the most widely used measures for genetic differentiation. F_{ST} ranges from 0.0 (no differentiation) to 1.0 (complete differentiation) (Ding et al., 2015).
3. The phenomenon of non-random co-occurrence of alleles at two loci on the same haplotype is known as linkage disequilibrium (LD) (Linkage disequilibrium could be caused by linkage (alleles are not passed independently to the next generation, which is especially likely for alleles located close to each other on the same chromosome, and population genetic effects such as population subdivision or admixture (Zaykin et al., 2008, Buckleton et al., 2005).
4. Polymorphism Information Content (PIC) of a marker corresponds to its ability to detect the polymorphism among individuals in a population. PIC values markers range from 0 (monomorphic) to 1 (very highly informative, with several alleles of equal frequency). According to Botstein et al. (1980), markers with PIC values greater than 0.5 are considered to be very informative, values between 0.25 and 0.50 are somewhat informative, and values lower than 0.25 are not very informative (Botstein et al., 1980; Serrote et al., 2020).
5. The paternity index (PI) reflects how many more times likely it is that the person being tested is the biological father, rather than a randomly selected individual. On the other hand, Power of exclusion (PE), is defined as the fraction of individuals having a DNA profile that is different from that of a randomly selected individual in a paternity case. (Huston, 1998).

$$PI \text{ or Typical paternity index (TPI)} = (H+h) / 2H$$

PE is calculated as: $PE = h^2(1-2hH^2)$

where h is the heterozygosity and H is the homozygosity at the locus.

6. Match Probability (PM) is defined as the probability of a match between two unrelated individuals. Power of Discrimination (PD) can be defined as the probability that two randomly selected individuals will have different genotypes (Goodwin et. al. 2011).

PM is calculated as:

$$\sum_{k=1}^m P k^2$$

Where, P_k is the frequency of each distinct genotype and m represents the number of the distinctive genotypes.

PD is calculated for a single locus =

$$1 - PM$$

For several loci $P_{D_{comb}} =$

$$1 - \prod_{i=1}^n [1 - P_{Di}]$$

Where, PM is the match probability of a single locus, $P_{D_{comb}}$ is the power of discrimination of several loci, P_{Di} is the power of discrimination of individual locus, and \prod stands for multiplication.

9.4 Appendix 4: Table describes the Ion Sphere™ particle (ISP) density, summary, and read length quality metrics for the unaligned sequencing reads.

Metric	Description
Total Bases	Number of filtered and trimmed base pairs reported in the output BAM file.
Key Signal	The reported key signal is the average signal for all ISPs that identically match the library key.
ISP Loading	Percentage of chip wells that contain a live ISP. (The percentage value considers only potentially addressable wells.)
Total Reads	Total number of filtered and trimmed reads independent of length reported in the output BAM file.
Usable Reads	The percentage of library ISPs that pass the polyclonal, low quality, and primer-dimer filters. This percentage is calculated by dividing final library ISPs by library ISPs.
Loading	Percentage of chip wells that contain a live ISP. (The percentage value considers only potentially addressable wells.) This percentage is calculated by dividing the number of loaded ISPs by the number of potentially addressable wells.
Empty Wells	Percentage of chip wells that do not contain an ISP. (The percentage value considers only potentially addressable wells.) The percentage is calculated by the number of potentially addressable wells, minus the number of loaded ISPs, divided by the number of potentially addressable wells.
Enrichment	Predicted number of live ISPs that have a key signal identical to the library key signal. The Percent Enrichment value reported is the number of loaded ISPs that are Library ISPs, after taking out Test Fragment (TF) ISPs. This number is calculated by dividing library ISPs by the number of loaded ISPs minus TF ISPs.
No Template	Percentage of chip wells that do not contain a DNA template. This percentage is calculated by the number of loaded ISPs minus TF ISPs divided by the number of loaded ISPs plus TF ISPs.

Metric	Description
Clonal	Percentage of clonal ISPs (all library and TF ISPs that are not polyclonal). An ISP is clonal if all of its DNA fragments are cloned from a single original template. All the fragments on such a bead are identical and they respond in unison as each nucleotide is flowed in turn across the chip. This percentage is calculated by dividing the number of ISPs with single beads by the number of live wells.
Polyclonal	Percentage of polyclonal ISPs (ISPs carrying clones from two or more templates). A high polyclonal percentage indicates that library input is too high and should be titrated down. Enrichment does not filter out polyclonal ISPs, it only removes template-negative ISPs. This percentage is calculated by dividing polyclonal ISPs by live ISPs.
Final Library	Percentage of reads that pass all filters and are recorded in the output BAM file. This value can be different from the Total Reads due to specifications associated with read trimming beyond a minimal requirement resulting in Total Reads being slightly less than Final Library. This percentage is calculated by dividing final library ISPs by clonal ISPs.
% Test Fragments	Percentage of live ISPs with a key signal that is identical to the TF key signal. This percentage is calculated by dividing TF ISPs by clonal ISPs.
% Adapter Dimer	Percentage of ISPs with an insert length of less than 8 bp. This percentage is calculated by dividing primer-dimer ISPs by clonal ISPs.
% Low Quality	Percentage of ISPs with a low or unrecognizable signal. This percentage is calculated by dividing low quality ISPs by clonal ISPs

9.5 Appendix 5: Data showing Heterozygote balance values as minimum, maximum and median Hb. The lowest median values are highlighted in yellow color.

SNP names	Minimum Hb	Maximum Hb	Median Hb
rs1490413	0.686975	0.991404	0.907749
rs7520386	0.154884	0.61583	0.404953
rs4847034	0.129562	0.993631	0.650343
rs560681	0.702918	1	0.929302
rs10495407	0.442308	0.99639	0.894835
rs891700	0.681672	0.99726	0.900356
rs1413212	0.55615	0.986301	0.821869
rs876724	0.385776	0.866667	0.565657
rs1109037	0.7675	1	0.929992
rs993934	0.327103	0.988095	0.773756
rs12997453	0.227273	1	0.832031
rs907100	0.670732	1	0.899123
rs1357617	0.590769	0.993377	0.913556
rs4364205	0.693966	0.995082	0.881935
rs1872575	0.736318	0.994382	0.901753
rs1355366	0.637168	0.996485	0.907395
rs6444724	0.621212	0.993548	0.844697
rs2046361	0.613445	0.989286	0.900431
rs6811238	0.509579	1	0.893773
rs1979255	0.820244	0.995763	0.928058
rs717302	0.243243	0.916667	0.473856
rs159606	0.372263	0.991549	0.778364
rs7704770	0.684211	0.995595	0.892694
rs251934	0.648649	1	0.94368
rs338882	0.751781	1	0.917282
rs13218440	0.656627	0.995227	0.87384
rs214955	0.361446	0.863636	0.645161
rs727811	0.614634	0.994681	0.911155
rs6955448	0.726829	1	0.917266

SNP names	Minimum Hb	Maximum Hb	Median Hb
rs917118	0.261905	0.647887	0.443662
rs321198	0.661491	1	0.862249
rs737681	0.59901	0.990909	0.913449
rs10092491	0.755507	0.998601	0.915888
rs4288409	0.592145	0.994186	0.903623
rs2056277	0.725389	1	0.911475
rs1015250	0.515625	0.985507	0.822165
rs7041158	0.793893	1	0.935323
rs1463729	0.723333	0.997475	0.893863
rs1360288	0.748214	0.990066	0.9
rs10776839	0.754875	1	0.886256
rs826472	0.132353	0.995305	0.75
rs735155	0.763441	1	0.928141
rs3780962	0.477178	0.960347	0.714432
rs740598	0.726154	0.998236	0.898077
rs964681	0.767313	0.996212	0.906593
rs1498553	0.677871	0.995575	0.912131
rs901398	0.691358	0.994536	0.918759
rs10488710	0.636364	1	0.885246
rs2076848	0.695652	1	0.889831
rs2269355	0.210312	1	0.918394
rs2111980	0.727273	0.97861	0.888036
rs10773760	0.656904	0.979104	0.863036
rs1335873	0.626761	0.996942	0.895758
rs1886510	0.709486	1	0.923547
rs1058083	0.73176	0.99679	0.918605
rs354439	0.669951	1	0.889447
rs1454361	0.734475	0.99845	0.92691
rs722290	0.643836	1	0.93913
rs873196	0.566154	0.987315	0.773333
rs4530059	0.744444	0.994975	0.890141
rs2016276	0.643192	0.972561	0.866142

SNP names	Minimum Hb	Maximum Hb	Median Hb
rs1821380	0.766798	0.998282	0.896552
rs1528460	0.685237	0.993644	0.931357
rs729172	0.641667	0.994792	0.934708
rs2342747	0.71831	1	0.857977
rs430046	0.586716	0.99569	0.860927
rs1382387	0.780083	0.996622	0.915015
rs9905977	0.636719	1	0.880383
rs740910	0.321127	1	0.915068
rs938283	0.7375	0.995058	0.907635
rs2292972	0.609375	0.996416	0.875241
rs1493232	0.429907	0.997468	0.831808
rs9951171	0.718487	1	0.913183
rs1736442	0.75378	0.99635	0.918089
rs1024116	0.64375	0.99854	0.879837
rs719366	0.502538	1	0.788764
rs576261	0.715152	0.994737	0.870536
rs1031825	0.415507	1	0.796216
rs445251	0.649321	0.996795	0.874489
rs1005533	0.523227	0.989041	0.910828
rs1523537	0.787589	0.995565	0.892159
rs722098	0.650997	1	0.912236
rs2830795	0.754098	0.993064	0.903077
rs2831700	0.466667	0.99154	0.815741
rs914165	0.69	0.997135	0.9
rs221956	0.702665	0.988889	0.913685
rs733164	0.741259	1	0.920732
rs987640	0.546185	1	0.849057
rs2040411	0.639423	1	0.901515
rs1028528	0.672535	1	0.895141

9.6 Appendix 6: Data showing noise level values as minimum, maximum and median of the Precision ID Identity panel. Y SNPs are shaded in grey.

SNP names	Noise Levels (NL)		
	Median	Minimum	Maximum
rs1490413	0.001577	0.000421	0.005714
rs7520386	0.0024	0.000259	0.100982
rs4847034	0.004178	0.0008	0.04878
rs560681	0.001908	0.000264	0.023301
rs10495407	0.002427	0.000779	0.04878
rs891700	0.001953	0.000663	0.014888
rs1413212	0.002646	0.000693	0.009569
rs876724	0.004878	0.000623	0.041667
rs1109037	0.001649	0.000416	0.005549
rs993934	0.003711	0.000998	0.015625
rs12997453	0.003333	0.001504	0.009615
rs907100	0.002322	0.000799	0.012295
rs1357617	0.002354	0.000678	0.025039
rs4364205	0.001797	0.0003	0.007972
rs1872575	0.001389	0.000501	0.006472
rs1355366	0.00175	0.000337	0.005294
rs6444724	0.003572	0.000694	0.009901
rs2046361	0.002857	0.000934	0.029851
rs6811238	0.001759	0.000528	0.007874
rs1979255	0.002177	0.000746	0.011136
rs717302	0.003835	0.000789	0.019608
rs159606	0.002188	0.000959	0.007782
rs7704770	0.002342	0.001062	0.009554
rs251934	0.001795	0.00033	0.009535
rs338882	0.001923	0.000605	0.007126
rs13218440	0.002646	0.000656	0.010811
rs214955	0.006757	0.003717	0.017544
rs727811	0.001353	0.000582	0.006917
rs6955448	0.001684	0.000429	0.003808
rs917118	0.003802	0.00122	0.013953
rs321198	0.002208	0.000473	0.063492
rs737681	0.00378	0.000485	0.081435
rs10092491	0.00425	0.000276	0.019934
rs4288409	0.001718	0.000645	0.004458
rs2056277	0.001901	0.000568	0.006116
rs1015250	0.003497	0.000917	0.006536
rs7041158	0.00213	0.00072	0.012987
rs1463729	0.002481	0.000524	0.061298
rs1360288	0.001086	0.000433	0.014265
rs10776839	0.003333	0.00057	0.076389
rs826472	0.002857	0.000868	0.116883
rs735155	0.001892	0.000524	0.009324
rs3780962	0.00304	0.000923	0.008

SNP names	Noise Levels (NL)		
	Median	Minimum	Maximum
rs740598	0.00211	0.000563	0.008475
rs964681	0.002415	0.000754	0.025641
rs1498553	0.002069	0.000305	0.00774
rs901398	0.003989	0.000385	0.018135
rs10488710	0.003478	0.000978	0.00916
rs2076848	0.002195	0.000418	0.008197
rs2269355	0.001394	0.000309	0.064246
rs2111980	0.002332	0.000584	0.005995
rs10773760	0.001985	0.000386	0.008772
rs1335873	0.002038	0.000379	0.00823
rs1886510	0.002053	0.000693	0.023256
rs1058083	0.001372	0.00023	0.007964
rs354439	0.002304	0.000423	0.016892
rs1454361	0.001825	0.000483	0.007905
rs722290	0.003115	0.001285	0.011268
rs873196	0.005398	0.001449	0.018898
rs4530059	0.00289	0.00043	0.009195
rs2016276	0.002029	0.000576	0.011129
rs1821380	0.001311	0.000313	0.017857
rs1528460	0.00166	0.000601	0.005952
rs729172	0.002222	0.000658	0.013245
rs2342747	0.004692	0.001661	0.010929
rs430046	0.001647	0.00059	0.004963
rs1382387	0.002485	0.000369	0.010331
rs9905977	0.001522	0.000348	0.008696
rs740910	0.002125	0.000556	0.041322
rs938283	0.00138	0.000377	0.008518
rs2292972	0.002424	0.000448	0.010714
rs1493232	0.005443	0.000668	0.023077
rs9951171	0.001996	0.000417	0.016722
rs1736442	0.001702	0.000491	0.007188
rs1024116	0.002849	0.000547	0.190476
rs719366	0.002326	0.000705	0.006466
rs576261	0.003067	0.000522	0.013495
rs1031825	0.004974	0.000331	0.026936
rs445251	0.001909	0.000488	0.008681
rs1005533	0.001366	0.000533	0.005666
rs1523537	0.00274	0.000822	0.064924
rs722098	0.001656	0.000335	0.007179
rs2830795	0.002	0.000512	0.005382
rs2831700	0.002511	0.000791	0.005698
rs914165	0.002227	0.000664	0.007435
rs221956	0.002036	0.000307	0.006737
rs733164	0.002177	0.001009	0.01
rs987640	0.005085	0.000965	0.052632
rs2040411	0.001887	0.000641	0.006623
rs1028528	0.001837	0.000341	0.007634
rs2534636	0.004435	0.001786	0.013043

SNP names	Noise Levels (NL)		
	Median	Minimum	Maximum
rs35284970	0.005118	0.002288	0.013483
rs9786184	0.00263	0.001185	0.010661
rs9786139	0.003049	0.001972	0.005076
rs16981290	0.00376	0.002481	0.009709
rs17250845	0.002849	0.001105	0.007859
L298	0.003681	0.00117	0.010526
P256	0.001956	0.001422	0.004739
P202	0.004065	0.001174	0.006993
rs17306671	0.003126	0.000873	0.015385
rs4141886	0.004808	0.002591	0.020833
rs2032595	0.00303	0.001695	0.009091
rs2032599	0.003639	0.001908	0.009901
rs20320	0.005051	0.001808	0.006969
rs2032602	0.003311	0.001063	0.03386
rs8179021	0.002725	0.001276	0.003953
rs2032624	0.006536	0.002703	0.020833
rs2032636	0.015275	0.002193	0.0983
rs9341278	0.006711	0.002119	0.027027
rs2032658	0.003287	0.001435	0.011429
rs2319818	0.004146	0.002075	0.004854
rs17269816	0.00287	0.00081	0.004566
rs17222573	0.003125	0.000838	0.008065
M479	0.016949	0.012658	0.071429
rs3848982	0.006061	0.001206	0.014218
rs3900	0.002747	0.000667	0.021858
rs3911	0.003652	0.000944	0.037037
rs2032631	0.006814	0.001597	0.159817
rs2032673	0.00225	0.000926	0.007092
rs2032652	0.003205	0.000756	0.007353
rs16980426	0.002436	0.00081	0.012903
rs13447443	0.006538	0.001299	0.022727
rs17842518	0.003682	0.001134	0.010363
rs2033003	0.00376	0.001007	0.016949

9.7 Appendix 7: Data showing allele frequency of 90 autosomal SNPs in Qatar population.

dbSNP_rs numbers	FREQ OF 'A'	FREQ OF 'C'	FREQ OF 'G'	FREQ OF 'T'
rs1490413	0.490		0.510	
rs7520386	0.581		0.419	
rs4847034	0.495		0.505	
rs560681	0.571		0.429	
rs10495407	0.362		0.638	
rs891700	0.395		0.605	
rs1413212		0.738		0.262
rs876724		0.757		0.243
rs1109037	0.390		0.610	
rs993934	0.624		0.376	
rs12997453	0.376		0.624	
rs907100		0.452	0.548	
rs1357617	0.290			0.710
rs4364205			0.595	0.405
rs1872575	0.581		0.419	
rs1355366		0.338		0.662
rs6444724		0.395		0.605
rs2046361	0.314			0.686
rs6811238			0.410	0.590
rs1979255		0.357	0.643	
rs717302	0.567		0.433	
rs159606	0.338		0.662	
rs7704770	0.643		0.357	
rs251934	0.557		0.443	
rs338882	0.471		0.529	
rs13218440	0.414		0.586	
rs214955		0.290		0.710
rs727811			0.524	0.476
rs6955448		0.710		0.290
rs917118		0.652		0.348
rs321198		0.629		0.371
rs737681		0.648		0.352
rs10092491		0.662		0.338
rs4288409	0.257	0.743		
rs2056277		0.767		0.233
rs1015250		0.233	0.767	
rs7041158		0.729		0.271
rs1463729		0.486		0.514
rs1360288		0.729		0.271
rs10776839			0.500	0.500
rs826472		0.643		0.357
rs735155		0.333		0.667
rs3780962	0.390		0.610	
rs740598	0.500		0.500	
rs964681		0.348		0.652

dbSNP_rs numbers	FREQ OF 'A'	FREQ OF 'C'	FREQ OF 'G'	FREQ OF 'T'
rs1498553		0.410		0.590
rs901398		0.243		0.757
rs10488710		0.500	0.500	
rs2076848	0.610			0.390
rs2269355		0.476	0.524	
rs2111980		0.390		0.610
rs10773760	0.619		0.381	
rs1335873	0.586			0.414
rs1886510	0.395		0.605	
rs1058083	0.386		0.614	
rs354439	0.538			0.462
rs1454361	0.443			0.557
rs722290		0.510	0.490	
rs873196		0.338		0.662
rs4530059	0.424		0.576	
rs2016276		0.129		0.871
rs1821380		0.638	0.362	
rs1528460		0.343		0.657
rs729172			0.614	0.386
rs2342747	0.286		0.714	
rs430046		0.614		0.386
rs1382387	0.790	0.210		
rs9905977	0.324		0.676	
rs740910	0.729		0.271	
rs938283		0.281		0.719
rs2292972		0.371		0.629
rs1493232	0.705	0.295		
rs9951171	0.395		0.605	
rs1736442		0.481		0.519
rs1024116		0.514		0.486
rs719366	0.738		0.262	
rs576261	0.624	0.376		
rs1031825	0.438	0.562		
rs445251		0.571	0.429	
rs1005533	0.433		0.567	
rs1523537		0.467		0.533
rs722098	0.733		0.267	
rs2830795	0.738		0.262	
rs2831700	0.552		0.448	
rs914165	0.500		0.500	
rs221956		0.629		0.371
rs733164	0.348		0.652	
rs987640	0.443			0.557
rs2040411	0.614		0.386	
rs1028528	0.614		0.386	

9.8 Appendix 8: Data showing forensic parameters of 90 autosomal SNPS of the Precision ID Identity panel for 105 Qatari samples, (PM: probability of match, PD: power of discrimination, PE: Power of Exclusion, PIC: polymorphism information content, and TPI: Typical Paternity Index).

#	SNP Names	PM	PIC	PD	PE	TPI
1	rs1490413	0.388	0.37	0.612	0.209	1.05
2	rs7520386	0.353	0.37	0.647	0.114	0.83
3	rs4847034	0.356	0.37	0.644	0.153	0.92
4	rs560681	0.355	0.37	0.645	0.126	0.86
5	rs10495407	0.382	0.36	0.618	0.126	0.86
6	rs891700	0.390	0.36	0.610	0.175	0.97
7	rs1413212	0.447	0.31	0.553	0.070	0.73
8	rs876724	0.466	0.30	0.534	0.070	0.73
9	rs1109037	0.380	0.36	0.620	0.153	0.92
10	rs993934	0.369	0.36	0.631	0.108	0.82
11	rs12997453	0.391	0.36	0.609	0.160	0.94
12	rs907100	0.351	0.37	0.649	0.132	0.88
13	rs1357617	0.426	0.33	0.574	0.108	0.82
14	rs4364205	0.417	0.37	0.583	0.228	1.09
15	rs1872575	0.377	0.37	0.623	0.168	0.95
16	rs1355366	0.405	0.35	0.595	0.146	0.91
17	rs6444724	0.364	0.36	0.636	0.120	0.85
18	rs2046361	0.404	0.34	0.596	0.066	0.72
19	rs6811238	0.380	0.37	0.620	0.168	0.95
20	rs1979255	0.376	0.35	0.624	0.097	0.80
21	rs717302	0.356	0.37	0.644	0.132	0.88
22	rs159606	0.405	0.35	0.595	0.146	0.91
23	rs7704770	0.409	0.35	0.591	0.175	0.97
24	rs251934	0.359	0.37	0.641	0.146	0.91
25	rs338882	0.389	0.37	0.611	0.209	1.05
26	rs13218440	0.392	0.37	0.608	0.192	1.01

#	SNP Names	PM	PIC	PD	PE	TPI
27	rs214955	0.430	0.33	0.570	0.120	0.85
28	rs727811	0.394	0.37	0.606	0.218	1.07
29	rs6955448	0.426	0.33	0.574	0.108	0.82
30	rs917118	0.388	0.35	0.612	0.120	0.85
31	rs321198	0.373	0.36	0.627	0.114	0.83
32	rs737681	0.380	0.35	0.620	0.103	0.81
33	rs10092491	0.399	0.35	0.601	0.132	0.88
34	rs4288409	0.458	0.31	0.542	0.114	0.83
35	rs2056277	0.476	0.29	0.524	0.078	0.75
36	rs1015250	0.476	0.29	0.524	0.070	0.73
37	rs7041158	0.440	0.32	0.560	0.097	0.80
38	rs1463729	0.350	0.37	0.650	0.139	0.89
39	rs1360288	0.457	0.32	0.543	0.146	0.91
40	rs10776839	0.342	0.38	0.658	0.120	0.85
41	rs826472	0.409	0.35	0.591	0.153	0.97
42	rs735155	0.392	0.35	0.608	0.103	0.81
43	rs3780962	0.380	0.36	0.620	0.153	0.92
44	rs740598	0.388	0.38	0.612	0.209	1.05
45	rs964681	0.393	0.35	0.607	0.132	0.88
46	rs1498553	0.351	0.37	0.649	0.092	0.78
47	rs901398	0.466	0.30	0.534	0.087	0.77
48	rs10488710	0.368	0.38	0.632	0.175	0.97
49	rs2076848	0.442	0.36	0.558	0.258	1.17
50	rs2269355	0.336	0.37	0.664	0.092	0.78
51	rs2111980	0.368	0.36	0.632	0.126	0.86
52	rs10773760	0.373	0.36	0.627	0.126	0.86
53	rs1335873	0.392	0.37	0.608	0.192	1.01
54	rs1886510	0.382	0.36	0.618	0.160	0.94
55	rs1058083	0.394	0.36	0.606	0.175	0.97
56	rs354439	0.371	0.37	0.629	0.175	0.97
57	rs1454361	0.384	0.37	0.616	0.192	1.01
58	rs722290	0.360	0.37	0.640	0.160	0.94

#	SNP Names	PM	PIC	PD	PE	TPI
59	rs873196	0.394	0.35	0.606	0.120	0.85
60	rs4530059	0.354	0.37	0.646	0.120	0.85
61	rs2016276	0.636	0.20	0.364	0.030	0.63
62	rs1821380	0.443	0.36	0.557	0.238	1.12
63	rs1528460	0.406	0.35	0.594	0.153	0.92
64	rs729172	0.373	0.36	0.627	0.132	0.88
65	rs2342747	0.429	0.32	0.571	0.103	0.81
66	rs430046	0.386	0.36	0.614	0.160	0.94
67	rs1382387	0.502	0.28	0.498	0.074	0.74
68	rs9905977	0.402	0.34	0.598	0.114	0.83
69	rs740910	0.447	0.32	0.553	0.120	0.85
70	rs938283	0.431	0.32	0.569	0.097	0.80
71	rs2292972	0.373	0.36	0.627	0.114	0.83
72	rs1493232	0.418	0.33	0.582	0.092	0.78
73	rs9951171	0.399	0.36	0.601	0.192	1.01
74	rs1736442	0.400	0.37	0.600	0.228	1.09
75	rs1024116	0.373	0.37	0.627	0.183	0.99
76	rs719366	0.449	0.31	0.551	0.097	0.80
77	rs576261	0.384	0.36	0.616	0.146	0.91
78	rs1031825	0.401	0.37	0.599	0.218	1.07
79	rs445251	0.345	0.37	0.655	0.092	0.78
80	rs1005533	0.369	0.37	0.631	0.160	0.94
81	rs1523537	0.352	0.37	0.648	0.139	0.89
82	rs722098	0.443	0.31	0.557	0.066	0.72
83	rs2830795	0.460	0.31	0.540	0.132	0.88
84	rs2831700	0.369	0.37	0.631	0.168	0.95
85	rs914165	0.368	0.38	0.632	0.175	0.97
86	rs221956	0.406	0.36	0.594	0.183	0.99
87	rs733164	0.399	0.35	0.601	0.146	0.91
88	rs987640	0.406	0.37	0.594	0.228	1.09
89	rs2040411	0.394	0.36	0.606	0.175	0.97
90	rs1028528	0.362	0.36	0.638	0.097	0.80

9.9 Appendix 9: Data showing observed [Obs.Het] and expected [Exp.Het] heterozygosities and P values from an exact test for HWE across 90 SNPs typed in 105 individuals from three populations. Green represents minimum Obs.Het and orange represents minimum Exp.Het.

SNP#	SNP ID	P1			P2			P3		
		Obs.Het	Exp.Het	p-value	Obs.Het	Exp.Het	p-value	Obs.Het	Exp.Het	p-value
1	rs1490413	0.41935	0.49974	0.46933	0.51429	0.50352	1	0.61538	0.5035	0.20374
2	rs7520386	0.41935	0.49974	0.46933	0.34286	0.50683	0.08952	0.41026	0.4662	0.49907
3	rs4847034	0.41935	0.50608	0.47027	0.45714	0.50352	0.73438	0.48718	0.50516	1
4	rs560681	0.54839	0.49974	0.71783	0.45714	0.49689	0.73332	0.28205	0.49451	0.0092
5	rs10495407	0.48387	0.45531	1	0.37143	0.50062	0.17297	0.41026	0.43157	1
6	rs891700	0.3871	0.50767	0.27943	0.51429	0.45714	0.70691	0.53846	0.47319	0.49851
7	rs1413212	0.29032	0.40455	0.1709	0.22857	0.35776	0.04753	0.41026	0.41026	1
8	rs876724	0.29032	0.33686	0.58199	0.34286	0.35776	1	0.30769	0.41026	0.13176
9	rs1109037	0.54839	0.48916	0.7081	0.37143	0.46584	0.27801	0.46154	0.49018	0.74744
10	rs993934	0.29032	0.48916	0.02992	0.42857	0.46584	0.7185	0.4359	0.47319	0.73343
11	rs12997453	0.48387	0.47435	1	0.34286	0.47371	0.14656	0.5641	0.47952	0.3193
12	rs907100	0.45161	0.50767	0.71944	0.37143	0.50559	0.17241	0.46154	0.49018	0.74912
13	rs1357617	0.41935	0.45531	0.70164	0.37143	0.30683	0.56842	0.38462	0.45854	0.47582
14	rs4364205	0.67742	0.5082	0.08017	0.34286	0.45714	0.15269	0.61538	0.47952	0.09626
15	rs1872575	0.41935	0.50608	0.46845	0.42857	0.46584	0.71844	0.5641	0.49817	0.51761
16	rs1355366	0.45161	0.48228	1	0.45714	0.45714	1	0.4359	0.42125	1
17	rs6444724	0.35484	0.45531	0.24742	0.4	0.50352	0.30645	0.46154	0.47952	1
18	rs2046361	0.41935	0.43205	1	0.22857	0.43727	0.00648	0.28205	0.44123	0.03115
19	rs6811238	0.54839	0.48916	0.70859	0.37143	0.50062	0.17262	0.51282	0.47952	0.74321
20	rs1979255	0.41935	0.48916	0.47227	0.4	0.49689	0.3078	0.30769	0.38628	0.22333
21	rs717302	0.3871	0.49498	0.27798	0.42857	0.50559	0.49821	0.46154	0.49018	0.74918
22	rs159606	0.41935	0.48916	0.47275	0.37143	0.42609	0.44839	0.53846	0.44123	0.26856
23	rs7704770	0.3871	0.44421	0.67897	0.54286	0.42609	0.12571	0.51282	0.49817	1

SNP#	SNP ID	P1			P2			P3		
		Obs.Het	Exp.Het	p-value	Obs.Het	Exp.Het	p-value	Obs.Het	Exp.Het	p-value
24	rs251934	0.32258	0.48228	0.12559	0.62857	0.50683	0.18632	0.38462	0.49451	0.19738
25	rs338882	0.3871	0.49498	0.27903	0.6	0.50062	0.30869	0.5641	0.49018	0.50652
26	rs13218440	0.58065	0.49498	0.46007	0.54286	0.48075	0.4932	0.41026	0.49817	0.3354
27	rs214955	0.48387	0.43205	0.67733	0.37143	0.42609	0.44899	0.38462	0.3986	1
28	rs727811	0.6129	0.49974	0.27802	0.4	0.49689	0.30637	0.58974	0.50117	0.33528
29	rs6955448	0.35484	0.40455	0.64915	0.48571	0.48075	1	0.33333	0.34532	1
30	rs917118	0.3871	0.48228	0.44562	0.37143	0.42609	0.44878	0.46154	0.4662	1
31	rs321198	0.35484	0.43205	0.40145	0.4	0.47371	0.46945	0.4359	0.49451	0.51973
32	rs737681	0.45161	0.49498	0.71812	0.2	0.40124	0.00553	0.48718	0.47319	1
33	rs10092491	0.45161	0.46536	1	0.4	0.41408	1	0.4359	0.47319	0.73431
34	rs4288409	0.41935	0.43205	1	0.25714	0.22733	1	0.51282	0.45022	0.47964
35	rs2056277	0.32258	0.31729	1	0.25714	0.37308	0.07955	0.41026	0.38628	1
36	rs1015250	0.22581	0.33686	0.09236	0.34286	0.2882	0.55833	0.35897	0.43157	0.44671
37	rs7041158	0.48387	0.45531	1	0.4	0.35776	0.65101	0.25641	0.38628	0.08154
38	rs1463729	0.45161	0.49498	0.71837	0.48571	0.46584	1	0.38462	0.48518	0.31168
39	rs1360288	0.41935	0.40455	1	0.54286	0.44762	0.26152	0.38462	0.34532	0.65498
40	rs10776839	0.41935	0.50608	0.47004	0.31429	0.50062	0.03932	0.48718	0.50516	1
41	rs826472	0.54839	0.47435	0.4567	0.45714	0.47371	1	0.46154	0.45022	1
42	rs735155	0.41935	0.45531	0.70025	0.37143	0.49234	0.1743	0.35897	0.38628	0.68477
43	rs3780962	0.35484	0.48916	0.14905	0.54286	0.46584	0.46133	0.46154	0.49018	0.74941
44	rs740598	0.51613	0.50767	1	0.51429	0.50352	1	0.53846	0.50117	0.74928
45	rs964681	0.58065	0.48228	0.28495	0.4	0.45714	0.47702	0.33333	0.44123	0.14973
46	rs1498553	0.51613	0.50344	1	0.31429	0.42609	0.21937	0.28205	0.48518	0.01671
47	rs901398	0.3871	0.35537	1	0.34286	0.41408	0.40592	0.33333	0.34532	1
48	rs10488710	0.54839	0.49974	0.71821	0.51429	0.50352	1	0.41026	0.50616	0.33738
49	rs2076848	0.67742	0.49974	0.06848	0.51429	0.48696	1	0.53846	0.45854	0.31205
50	rs2269355	0.45161	0.50344	0.71761	0.37143	0.50062	0.17281	0.28205	0.49451	0.00937
51	rs2111980	0.48387	0.48916	1	0.4	0.47371	0.47194	0.38462	0.48518	0.31236

SNP#	SNP ID	P1			P2			P3		
		Obs.Het	Exp.Het	p-value	Obs.Het	Exp.Het	p-value	Obs.Het	Exp.Het	p-value
52	rs10773760	0.51613	0.48228	1	0.4	0.47371	0.47063	0.35897	0.47952	0.17372
53	rs1335873	0.48387	0.49974	1	0.6	0.50559	0.3183	0.4359	0.45854	1
54	rs1886510	0.48387	0.49974	1	0.54286	0.50062	0.73506	0.38462	0.44123	0.47265
55	rs1058083	0.45161	0.49498	0.71847	0.42857	0.42609	1	0.5641	0.49817	0.51633
56	rs354439	0.45161	0.50767	0.71952	0.57143	0.50352	0.50217	0.4359	0.49451	0.51938
57	rs1454361	0.51613	0.49498	1	0.54286	0.50559	0.7404	0.46154	0.49817	0.74709
58	rs722290	0.45161	0.50344	0.71768	0.51429	0.50683	1	0.4359	0.50117	0.51956
59	rs873196	0.51613	0.46536	0.69855	0.37143	0.46584	0.27767	0.35897	0.43157	0.44762
60	rs4530059	0.3871	0.44421	0.67991	0.4	0.50683	0.30697	0.4359	0.50117	0.51897
61	rs2016276	0.19355	0.31729	0.05451	0.11429	0.10932	1	0.28205	0.24542	1
62	rs1821380	0.67742	0.47435	0.02129	0.45714	0.45714	1	0.53846	0.47319	0.49664
63	rs1528460	0.51613	0.46536	0.69893	0.34286	0.41408	0.40751	0.51282	0.47952	0.74239
64	rs729172	0.45161	0.48228	1	0.25714	0.44762	0.01867	0.5641	0.49817	0.51834
65	rs2342747	0.54839	0.45531	0.41966	0.22857	0.35776	0.04769	0.38462	0.42125	0.70199
66	rs430046	0.3871	0.46536	0.43689	0.51429	0.50352	1	0.48718	0.45854	0.73841
67	rs1382387	0.32258	0.27499	0.5683	0.31429	0.37308	0.37424	0.33333	0.34532	1
68	rs9905977	0.45161	0.46536	1	0.25714	0.42609	0.03697	0.48718	0.44123	0.71424
69	rs740910	0.51613	0.44421	0.43442	0.42857	0.40124	1	0.30769	0.35964	0.38369
70	rs938283	0.29032	0.43205	0.093	0.45714	0.41408	0.68824	0.35897	0.38628	0.68457
71	rs2292972	0.32258	0.46536	0.11845	0.51429	0.50683	1	0.35897	0.41026	0.44907
72	rs1493232	0.35484	0.43205	0.4026	0.4	0.41408	1	0.33333	0.42125	0.24995
73	rs9951171	0.48387	0.48916	1	0.6	0.49234	0.29466	0.4359	0.47319	0.73421
74	rs1736442	0.51613	0.50767	1	0.6	0.49234	0.29674	0.51282	0.5035	1
75	rs1024116	0.48387	0.49974	1	0.57143	0.49689	0.49016	0.4359	0.49451	0.52046
76	rs719366	0.48387	0.47435	1	0.31429	0.34161	0.63007	0.33333	0.34532	1
77	rs576261	0.48387	0.50608	1	0.54286	0.46584	0.46296	0.33333	0.3986	0.41157
78	rs1031825	0.3871	0.48228	0.44697	0.71429	0.49234	0.01253	0.48718	0.50649	1
79	rs445251	0.29032	0.47435	0.0505	0.31429	0.50725	0.04087	0.46154	0.49018	0.74909

SNP#	SNP ID	P1			P2			P3		
		Obs.Het	Exp.Het	p-value	Obs.Het	Exp.Het	p-value	Obs.Het	Exp.Het	p-value
80	rs1005533	0.48387	0.47435	1	0.37143	0.48075	0.28095	0.53846	0.50516	0.75304
81	rs1523537	0.3871	0.49498	0.27535	0.4	0.50352	0.30544	0.51282	0.49817	1
82	rs722098	0.32258	0.31729	1	0.22857	0.35776	0.04682	0.35897	0.4662	0.17531
83	rs2830795	0.41935	0.37282	0.64554	0.57143	0.45714	0.25493	0.30769	0.33034	0.64043
84	rs2831700	0.35484	0.48916	0.14813	0.62857	0.50352	0.17836	0.4359	0.48518	0.73681
85	rs914165	0.51613	0.50344	1	0.51429	0.49689	1	0.4359	0.50516	0.52059
86	rs221956	0.45161	0.49498	0.7185	0.62857	0.47371	0.07197	0.41026	0.45022	0.71783
87	rs733164	0.32258	0.35537	0.62248	0.48571	0.48075	1	0.51282	0.49018	1
88	rs987640	0.6129	0.48916	0.2597	0.42857	0.48075	0.72142	0.58974	0.50516	0.34483
89	rs2040411	0.51613	0.48228	1	0.51429	0.48696	1	0.4359	0.47319	0.73393
90	rs1028528	0.3871	0.50344	0.277	0.37143	0.46584	0.27865	0.35897	0.4662	0.17467

9.10 Appendix 10: Significant LD detected in P1, P2, P3.

9.10.1 Appendix A: Significant LD detected in P1 in yellow color.

9.10.2 Appendix B: Significant LD detected in P2 in orange color.

9.10.3 Appendix C: Significant LD detected in P3 in blue color.

9.11 Appendix 11: Data showing allele frequency of 165 autosomal SNPs in Qatar population.

SNP#	SNP ID	FREQ OF 'A'	FREQ OF 'C'	FREQ OF 'G'	FREQ OF 'T'
1	rs2986742		0.263		0.737
2	rs6541030	0.140		0.860	
3	rs647325	0.710		0.290	
4	rs4908343	0.563		0.437	
5	rs1325502	0.153		0.847	
6	rs12130799	0.986		0.014	
7	rs3118378	0.737		0.263	
8	rs3737576		0.033		0.967
9	rs7554936		0.316		0.684
10	rs2814778		0.364		0.636
11	rs1040404	0.488		0.512	
12	rs1407434	0.113		0.887	
13	rs4951629		0.048		0.952
14	rs316873		0.852		0.148
15	rs798443	0.692		0.308	
16	rs7421394	0.695		0.305	
17	rs1876482	0.034		0.966	
18	rs1834619	0.050		0.950	
19	rs4666200	0.619		0.381	
20	rs4670767			0.911	0.089
21	rs13400937			0.393	0.607
22	rs3827760	0.990		0.010	
23	rs260690	0.800	0.200		
24	rs6754311		0.967		0.033
25	rs10496971			0.159	0.841
26	rs10497191		0.827		0.173
27	rs2627037	0.107		0.893	
28	rs1569175		0.876		0.124
29	rs4955316			0.056	0.944
30	rs9809104		0.271		0.729
31	rs6548616		0.410		0.590
32	rs12629908	0.058		0.942	
33	rs12498138	0.084		0.916	
34	rs9845457	0.519		0.481	
35	rs734873	0.062		0.938	
36	rs2030763	0.198		0.802	
37	rs1513181	0.185		0.815	
38	rs9291090	0.916	0.084		
39	rs4833103	0.067	0.933		
40	rs10007810	0.304		0.696	
41	rs1369093		0.075		0.925
42	rs385194	0.448		0.552	
43	rs1229984		0.888		0.112
44	rs3811801			1.000	
45	rs7657799			0.091	0.909
46	rs2702414	0.073		0.927	
47	rs316598		0.497		0.503
48	rs870347	0.935	0.065		

SNP#	SNP ID	FREQ OF 'A'	FREQ OF 'C'	FREQ OF 'G'	FREQ OF 'T'
49	rs16891982		0.928	0.072	
50	rs37369		0.830		0.170
51	rs6451722	0.314		0.686	
52	rs12657828	0.829		0.171	
53	rs6556352		0.591		0.409
54	rs1500127		0.922		0.078
55	rs7722456	0.005	0.209		0.786
56	rs6422347		0.197		0.803
57	rs1040045	0.732		0.268	
58	rs2504853		0.336		0.664
59	rs7745461	0.220		0.780	
60	rs192655	0.883		0.117	
61	rs3823159	0.922		0.078	
62	rs4463276	0.476		0.524	
63	rs4458655		0.184		0.816
64	rs1871428	0.507		0.493	
65	rs731257	0.096		0.904	
66	rs917115		0.302		0.698
67	rs32314		0.216		0.784
68	rs2330442	0.577		0.423	
69	rs4717865	0.119		0.881	
70	rs10954737		0.038		0.962
71	rs705308	0.319	0.681		
72	rs7803075	0.483		0.517	
73	rs10236187	0.915	0.085		
74	rs6464211		0.710		0.290
75	rs10108270	0.425	0.575		
76	rs3943253	0.920		0.080	
77	rs1471939		0.253		0.747
78	rs1462906		0.847		0.153
79	rs12544346	0.380		0.620	
80	rs6990312			0.651	0.349
81	rs2196051	0.473		0.527	
82	rs7844723		0.646		0.354
83	rs2001907		0.958		0.042
84	rs1871534		0.080	0.920	
85	rs10511828		0.069		0.931
86	rs3793451		0.947		0.053
87	rs2306040		0.042		0.958
88	rs10513300		0.058		0.942
89	rs3814134	0.841		0.159	
90	rs2073821		0.881		0.119
91	rs3793791		0.114		0.886
92	rs4746136	0.075		0.925	
93	rs4918664	0.945		0.055	
94	rs4918842		0.130		0.870
95	rs4880436		0.979		0.021
96	rs10839880		0.396		0.604
97	rs1837606		0.278		0.722
98	rs2946788			0.198	0.802

SNP#	SNP ID	FREQ OF 'A'	FREQ OF 'C'	FREQ OF 'G'	FREQ OF 'T'
99	rs174570		0.920		0.080
100	rs11227699	0.040		0.960	
101	rs1079597		0.899		0.101
102	rs948028	0.712	0.288		
103	rs2416791	0.156		0.844	
104	rs1513056	0.127		0.873	
105	rs214678		0.129		0.871
106	rs772262	0.239		0.761	
107	rs2070586	0.232		0.768	
108	rs2238151		0.630		0.370
109	rs671			1.000	
110	rs1503767			0.266	0.734
111	rs9319336		0.068		0.932
112	rs7997709		0.138		0.862
113	rs1572018		0.663		0.337
114	rs2166624	0.312		0.688	
115	rs7326934		0.140	0.860	
116	rs9530435		0.652		0.348
117	rs9522149		0.671		0.329
118	rs1760921		0.142		0.858
119	rs2357442	0.774	0.226		
120	rs1950993			0.509	0.491
121	rs8021730			0.730	0.270
122	rs946918			0.824	0.176
123	rs200354			0.846	0.154
124	rs3784230	0.589		0.411	
125	rs1800414				1.000
126	rs12913832	0.957		0.043	
127	rs12439433	0.991		0.009	
128	rs735480		0.156		0.844
129	rs1426654	0.918		0.082	
130	rs2899826	0.853		0.147	
131	rs8035124	0.634	0.364	0.002	
132	rs4984913	0.700		0.300	
133	rs4781011			0.714	0.286
134	rs818386		0.942		0.058
135	rs2966849	0.072		0.928	
136	rs459920		0.505		0.495
137	rs1879488	0.043	0.957		
138	rs4411548		0.862		0.138
139	rs2593595	0.684		0.316	
140	rs17642714	0.836			0.164
141	rs4471745	0.011		0.989	
142	rs2033111	0.778		0.222	
143	rs11652805		0.395		0.605
144	rs10512572	0.100		0.900	
145	rs2125345		0.542		0.458
146	rs4798812	0.485		0.515	
147	rs2042762		0.005		0.995
148	rs7226659			0.888	0.112
149	rs7238445	0.345		0.655	
150	rs881728	0.181	0.819		

SNP#	SNP ID	FREQ OF 'A'	FREQ OF 'C'	FREQ OF 'G'	FREQ OF 'T'
151	rs3916235		0.862		0.138
152	rs4891825	0.841		0.159	
153	rs874299		0.705		0.295
154	rs7251928	0.643	0.357		
155	rs8113143	0.618	0.382		
156	rs3745099	0.674		0.326	
157	rs2532060		0.624		0.376
158	rs6104567			0.215	0.785
159	rs3907047		0.022		0.978
160	rs310644		0.138		0.862
161	rs2835370		0.145		0.855
162	rs1296819	0.304	0.696		
163	rs4821004		0.566		0.434
164	rs2024566	0.690		0.310	
165	rs5768007		0.871		0.129

9.12 Appendix 12: Data showing the Observed (Obs.Het) and expected (Exp.Het) heterozygosities and p-values from an exact test for Hardy-Weinberg Equilibrium (HWE) across 165 SNPs. Cells highlighted with yellow represents *p-value <0.05, blue represents monomorphic SNPs.

SNP#	SNP ID	Obs.Het	Exp.Het	p-value
1	rs2986742	0.35417	0.38918	0.13459
2	rs6541030	0.25839	0.24014	0.23029
3	rs647325	0.38095	0.41176	0.20393
4	rs4908343	0.44369	0.49287	0.09750
5	rs1325502	0.24150	0.25734	0.35749
6	rs12130799	0.02807	0.02772	1.00000
7	rs3118378	0.36735	0.38728	0.36737
8	rs3737576	0.05842	0.06327	0.26371
9	rs7554936	0.41892	0.43168	0.68457
10	rs2814778	0.40690	0.46463	0.04116*
11	rs1040404	0.51014	0.50057	0.81455
12	rs1407434	0.19858	0.20155	0.76661
13	rs4951629	0.08276	0.09205	0.13380
14	rs316873	0.24828	0.25302	0.81492
15	rs798443	0.44257	0.42785	0.59107
16	rs7421394	0.38514	0.42393	0.13310
17	rs1876482	0.06122	0.06583	0.28489
18	rs1834619	0.09375	0.09579	0.52430
19	rs4666200	0.45973	0.47241	0.71158
20	rs4670767	0.14591	0.15946	0.14310
21	rs13400937	0.44286	0.47789	0.26076
22	rs3827760	0.01761	0.01748	1.00000
23	rs260690	0.30662	0.31888	0.57569
24	rs6754311	0.06711	0.06497	1.00000
25	rs10496971	0.28571	0.26907	0.38725
26	rs10497191	0.28620	0.28715	1.00000
27	rs2627037	0.18771	0.19223	0.75634
28	rs1569175	0.22148	0.21786	1.00000
29	rs4955316	0.10544	0.10613	1.00000
30	rs9809104	0.35473	0.39667	0.07825
31	rs6548616	0.48123	0.48508	0.90550
32	rs12629908	0.10993	0.11037	1.00000
33	rs12498138	0.15385	0.15402	1.00000
34	rs9845457	0.55442	0.50015	0.07947
35	rs734873	0.11111	0.11739	0.30041
36	rs2030763	0.32534	0.31682	0.71307
37	rs1513181	0.30976	0.30229	0.84696
38	rs9291090	0.16140	0.15451	0.70430
39	rs4833103	0.12795	0.12582	1.00000
40	rs10007810	0.40278	0.42512	0.40659
41	rs1369093	0.12892	0.13884	0.19965
42	rs385194	0.46622	0.49570	0.34789
43	rs1229984	0.18519	0.20048	0.23703
44	rs3811801	This locus is monomorphic		
45	rs7657799	0.15917	0.16403	0.71393

SNP#	SNP ID	Obs.Het	Exp.Het	p-value
46	rs2702414	0.13889	0.13543	1.00000
47	rs316598	0.50505	0.50079	0.90832
48	rs870347	0.13074	0.12241	0.61778
49	rs16891982	0.13058	0.13414	0.64910
50	rs37369	0.28125	0.28057	1.00000
51	rs6451722	0.46048	0.43059	0.27669
52	rs12657828	0.30928	0.28509	0.21225
53	rs6556352	0.46102	0.48344	0.46955
54	rs1500127	0.15331	0.14181	0.38802
55	rs7722456	0.32215	0.33996	0.45542
56	rs6422347	0.31142	0.31721	0.71386
57	rs1040045	0.34680	0.39271	0.05409
58	rs2504853	0.40678	0.44670	0.14962
59	rs7745461	0.31987	0.34059	0.30575
60	rs192655	0.21886	0.20568	0.39608
61	rs3823159	0.12287	0.14492	0.02214*
62	rs4463276	0.51014	0.49988	0.72928
63	rs4458655	0.28276	0.29923	0.33197
64	rs1871428	0.49320	0.50071	0.81703
65	rs731257	0.18151	0.17091	0.48912
66	rs917115	0.45085	0.42341	0.27651
67	rs32314	0.34471	0.33619	0.72860
68	rs2330442	0.48485	0.48830	0.90472
69	rs4717865	0.21724	0.20999	0.77859
70	rs10954737	0.07560	0.07287	1.00000
71	rs705308	0.36301	0.43486	0.00668*
72	rs7803075	0.48311	0.50015	0.56239
73	rs10236187	0.15068	0.15684	0.45232
74	rs6464211	0.43253	0.41161	0.47180
75	rs10108270	0.53020	0.48992	0.18964
76	rs3943253	0.13287	0.14816	0.09452
77	rs1471939	0.37884	0.37988	1.00000
78	rs1462906	0.23569	0.25755	0.17083
79	rs12544346	0.45085	0.47265	0.45826
80	rs6990312	0.44876	0.45463	0.89424
81	rs2196051	0.47099	0.49954	0.34872
82	rs7844723	0.45392	0.45771	0.89855
83	rs2001907	0.06944	0.08000	0.07826
84	rs1871534	0.12892	0.15062	0.02946*
85	rs10511828	0.13194	0.12947	1.00000
86	rs3793451	0.09898	0.10038	0.56770
87	rs2306040	0.07018	0.08081	0.07895
88	rs10513300	0.10959	0.10985	1.00000
89	rs3814134	0.22917	0.26889	0.02472*
90	rs2073821	0.22414	0.20999	0.39587
91	rs3793791	0.18182	0.20309	0.08206
92	rs4746136	0.13732	0.14019	0.66665
93	rs4918664	0.10638	0.10090	1.00000
94	rs4918842	0.23368	0.22745	0.79804
95	rs4880436	0.03546	0.04172	0.11141
96	rs10839880	0.47297	0.47887	0.90354
97	rs1837606	0.40268	0.40257	1.00000
98	rs2946788	0.32323	0.31892	1.00000

SNP#	SNP ID	Obs.Het	Exp.Het	p-value
99	rs174570	0.13014	0.14537	0.08734
100	rs11227699	0.08099	0.07784	1.00000
101	rs1079597	0.16151	0.18251	0.09599
102	rs948028	0.42177	0.40886	0.66811
103	rs2416791	0.24579	0.26455	0.26974
104	rs1513056	0.21951	0.22239	0.79095
105	rs214678	0.23860	0.22634	0.43992
106	rs772262	0.35531	0.36539	0.62184
107	rs2070586	0.35517	0.35778	0.87042
108	rs2238151	0.45270	0.46783	0.61925
109	rs671	This locus is monomorphic		
110	rs1503767	0.41053	0.39015	0.44662
111	rs9319336	0.13014	0.12782	1.00000
112	rs7997709	0.23569	0.23838	0.80809
113	rs1572018	0.43197	0.44856	0.60460
114	rs2166624	0.43151	0.42978	1.00000
115	rs7326934	0.23050	0.24133	0.45668
116	rs9530435	0.40940	0.45516	0.09501
117	rs9522149	0.40940	0.44216	0.23812
118	rs1760921	0.23944	0.24244	0.80682
119	rs2357442	0.35836	0.35150	0.86796
120	rs1950993	0.47312	0.50074	0.40609
121	rs8021730	0.36014	0.39256	0.17314
122	rs946918	0.29795	0.29103	0.84048
123	rs200354	0.20678	0.26134	0.00116*
124	rs3784230	0.48797	0.48425	0.90413
125	rs1800414	This locus is monomorphic		
126	rs12913832	0.07192	0.08209	0.08735
127	rs12439433	0.01761	0.01748	1.00000
128	rs735480	0.23129	0.26441	0.04395*
129	rs1426654	0.12195	0.15062	0.00618*
130	rs2899826	0.24742	0.24742	1.00000
131	rs8035124	0.51195	0.46661	0.05821
132	rs4984913	0.39726	0.42181	0.33150
133	rs4781011	0.38591	0.40988	0.32311
134	rs818386	0.11037	0.11039	1.00000
135	rs2966849	0.13758	0.13411	1.00000
136	rs459920	0.47651	0.50082	0.42077
137	rs1879488	0.08725	0.08358	1.00000
138	rs4411548	0.22727	0.23849	0.45142
139	rs2593595	0.46233	0.43107	0.22218
140	rs17642714	0.28082	0.27519	0.83315
141	rs4471745	0.02128	0.02109	1.00000
142	rs2033111	0.34014	0.34500	0.86568
143	rs11652805	0.41638	0.47915	0.03008*
144	rs10512572	0.18728	0.18145	0.75245
145	rs2125345	0.46284	0.49727	0.24651
146	rs4798812	0.48276	0.50048	0.56397
147	rs2042762	0.01087	0.01083	1.00000
148	rs7226659	0.17647	0.19997	0.06797
149	rs7238445	0.43051	0.45214	0.44100
150	rs881728	0.29066	0.29562	0.84192
151	rs3916235	0.22973	0.23906	0.46963

SNP#	SNP ID	Obs.Het	Exp.Het	p-value
152	rs4891825	0.26599	0.26916	0.82935
153	rs874299	0.39799	0.41746	0.48538
154	rs7251928	0.36552	0.46082	0.00042*
155	rs8113143	0.49662	0.47363	0.46301
156	rs3745099	0.41751	0.44061	0.42536
157	rs2532060	0.45638	0.46996	0.62089
158	rs6104567	0.34138	0.33872	1.00000
159	rs3907047	0.04407	0.04317	1.00000
160	rs310644	0.21818	0.23861	0.20048
161	rs2835370	0.27181	0.24975	0.16182
162	rs1296819	0.45000	0.42075	0.25755
163	rs4821004	0.47099	0.49198	0.48091
164	rs2024566	0.44983	0.42962	0.49373
165	rs5768007	0.21631	0.22576	0.43427

9.13 Appendix 13. Data showing the forensic parameters of 165 SNPS of The Precision ID Ancestry panel for 300 Qatari samples. PM: match probability, PD: power of discrimination, PE: Power of Exclusion, PIC: polymorphism information content and TPI: Typical Paternity Index). Cells highlighted with green color represents the highest discrimination power (PD) whereas cells highlighted with blue color represents the highest power of exclusion (PE).

SNP#	SNP ID	PM	PIC	PD	PE	TPI
1	rs2986742	0.446	0.31	0.554	0.088	0.77
2	rs6541030	0.600	0.21	0.400	0.049	0.68
3	rs647325	0.425	0.33	0.575	0.104	0.81
4	rs4908343	0.360	0.37	0.640	0.144	0.90
5	rs1325502	0.587	0.23	0.413	0.043	0.66
6	rs12130799	0.946	0.03	0.054	0.001	0.51
7	rs3118378	0.448	0.31	0.552	0.096	0.79
8	rs3737576	0.884	0.06	0.116	0.003	0.53
9	rs7554936	0.411	0.34	0.589	0.125	0.86
10	rs2814778	0.378	0.36	0.622	0.117	0.84
11	rs1040404	0.381	0.37	0.619	0.375	1.02
12	rs1407434	0.660	0.18	0.340	0.029	0.62
13	rs4951629	0.836	0.09	0.164	0.006	0.54
14	rs316873	0.593	0.22	0.407	0.044	0.66
15	rs798443	0.424	0.34	0.576	0.141	0.89
16	rs7421394	0.414	0.33	0.586	0.106	0.82
17	rs1876482	0.879	0.06	0.121	0.003	0.53
18	rs1834619	0.824	0.09	0.176	0.007	0.55
19	rs4666200	0.386	0.36	0.614	0.156	0.93
20	rs4670767	0.723	0.15	0.277	0.017	0.59
21	rs13400937	0.375	0.36	0.625	0.144	0.90
22	rs3827760	0.959	0.02	0.041	0.000	0.51
23	rs260690	0.515	0.27	0.485	0.067	0.72
24	rs6754311	0.875	0.06	0.125	0.004	0.54
25	rs10496971	0.569	0.23	0.431	0.057	0.70
26	rs10497191	0.551	0.25	0.449	0.058	0.70
27	rs2627037	0.674	0.17	0.326	0.026	0.62
28	rs1569175	0.635	0.19	0.365	0.036	0.64
29	rs4955316	0.806	0.10	0.194	0.009	0.56
30	rs9809104	0.439	0.32	0.561	0.088	0.77
31	rs6548616	0.382	0.37	0.618	0.170	0.96
32	rs12629908	0.799	0.10	0.201	0.010	0.56
33	rs12498138	0.729	0.14	0.271	0.018	0.59
34	rs9845457	0.408	0.37	0.592	0.241	1.13
35	rs734873	0.791	0.11	0.209	0.010	0.56
36	rs2030763	0.516	0.27	0.484	0.076	0.74
37	rs1513181	0.533	0.26	0.467	0.067	0.72
38	rs9291090	0.724	0.14	0.276	0.020	0.60
39	rs4833103	0.772	0.12	0.228	0.013	0.57
40	rs10007810	0.417	0.33	0.583	0.026	0.62
41	rs1369093	0.758	0.13	0.242	0.013	0.57
42	rs385194	0.365	0.37	0.635	0.158	0.93

SNP#	SNP ID	PM	PIC	PD	PE	TPI
43	rs1229984	0.667	0.18	0.333	0.026	0.61
44	rs3811801	1.000	0.00	0.000	0.000	0.50
45	rs7657799	0.711	0.15	0.289	0.020	0.60
46	rs2702414	0.756	0.13	0.244	0.015	0.58
47	rs316598	0.377	0.37	0.623	0.190	1.01
48	rs870347	0.773	0.11	0.227	0.014	0.57
49	rs16891982	0.762	0.12	0.238	0.014	0.57
50	rs37369	0.555	0.24	0.445	0.057	0.70
51	rs6451722	0.428	0.34	0.572	0.158	0.93
52	rs12657828	0.550	0.24	0.450	0.067	0.72
53	rs6556352	0.374	0.37	0.626	0.154	0.93
54	rs1500127	0.736	0.13	0.264	0.019	0.59
55	rs7722456	0.491	0.28	0.509	0.073	0.74
56	rs6422347	0.518	0.27	0.482	0.068	0.73
57	rs1040045	0.441	0.32	0.559	0.086	0.77
58	rs2504853	0.396	0.35	0.604	0.119	0.85
59	rs7745461	0.491	0.28	0.509	0.072	0.73
60	rs192655	0.645	0.19	0.355	0.036	0.64
61	rs3823159	0.756	0.13	0.244	0.012	0.57
62	rs4463276	0.380	0.37	0.620	0.195	1.02
63	rs4458655	0.537	0.26	0.463	0.058	0.70
64	rs1871428	0.375	0.37	0.629	0.180	0.98
65	rs731257	0.694	0.16	0.306	0.026	0.61
66	rs917115	0.432	0.33	0.568	0.147	0.91
67	rs32314	0.495	0.28	0.505	0.083	0.76
68	rs2330442	0.379	0.37	0.621	0.173	0.97
69	rs4717865	0.645	0.19	0.355	0.034	0.64
70	rs10954737	0.861	0.07	0.139	0.005	0.54
71	rs705308	0.400	0.34	0.600	0.094	0.79
72	rs7803075	0.367	0.37	0.633	0.172	0.96
73	rs10236187	0.728	0.14	0.272	0.018	0.59
74	rs6464211	0.437	0.33	0.563	0.136	0.88
75	rs10108270	0.402	0.37	0.598	0.214	1.06
76	rs3943253	0.746	0.14	0.254	0.014	0.58
77	rs1471939	0.458	0.31	0.542	0.101	0.80
78	rs1462906	0.588	0.23	0.412	0.041	0.66
79	rs12544346	0.382	0.36	0.618	0.147	0.91
80	rs6990312	0.400	0.35	0.600	0.148	0.91
81	rs2196051	0.363	0.37	0.637	0.162	0.94
82	rs7844723	0.399	0.35	0.601	0.152	0.92
83	rs2001907	0.858	0.08	0.142	0.004	0.54
84	rs1871534	0.751	0.14	0.249	0.013	0.57
85	rs10511828	0.766	0.12	0.234	0.014	0.58
86	rs3793451	0.816	0.09	0.184	0.008	0.55
87	rs2306040	0.857	0.08	0.143	0.004	0.54
88	rs10513300	0.799	0.10	0.201	0.010	0.56
89	rs3814134	0.582	0.23	0.418	0.038	0.65
90	rs2073821	0.642	0.19	0.358	0.036	0.64
91	rs3793791	0.666	0.18	0.334	0.025	0.61
92	rs4746136	0.752	0.13	0.248	0.015	0.58
93	rs4918664	0.805	0.10	0.195	0.010	0.56
94	rs4918842	0.622	0.20	0.378	0.039	0.65
95	rs4880436	0.925	0.04	0.075	0.001	0.52

SNP#	SNP ID	PM	PIC	PD	PE	TPI
96	rs10839880	0.385	0.36	0.615	0.166	0.95
97	rs1837606	0.439	0.32	0.561	0.115	0.84
98	rs2946788	0.516	0.27	0.484	0.073	0.74
99	rs174570	0.746	0.14	0.254	0.014	0.58
100	rs11227699	0.852	0.07	0.148	0.006	0.54
101	rs1079597	0.696	0.17	0.304	0.020	0.60
102	rs948028	0.434	0.33	0.566	0.127	0.86
103	rs2416791	0.582	0.23	0.418	0.043	0.66
104	rs1513056	0.632	0.20	0.368	0.035	0.64
105	rs214678	0.622	0.20	0.378	0.041	0.66
106	rs772262	0.470	0.30	0.530	0.088	0.77
107	rs2070586	0.478	0.29	0.522	0.088	0.77
108	rs2238151	0.388	0.36	0.612	0.148	0.91
109	rs671	1.000	0.00	0.000	0.000	0.50
110	rs1503767	0.453	0.31	0.547	0.122	0.85
111	rs9319336	0.768	0.12	0.232	0.014	0.57
112	rs7997709	0.611	0.21	0.389	0.040	0.65
113	rs1572018	0.400	0.35	0.600	0.134	0.88
114	rs2166624	0.419	0.34	0.581	0.136	0.88
115	rs7326934	0.609	0.21	0.391	0.038	0.65
116	rs9530435	0.388	0.35	0.612	0.119	0.84
117	rs9522149	0.401	0.34	0.599	0.121	0.85
118	rs1760921	0.602	0.21	0.398	0.042	0.66
119	rs2357442	0.484	0.29	0.516	0.090	0.78
120	rs1950993	0.364	0.37	0.636	0.167	0.95
121	rs8021730	0.440	0.32	0.560	0.091	0.78
122	rs946918	0.546	0.25	0.454	0.062	0.71
123	rs200354	0.597	0.23	0.403	0.031	0.63
124	rs3784230	0.384	0.37	0.616	0.176	0.97
125	rs1800414	1.000	0.00	0.000	0.000	0.50
126	rs12913832	0.854	0.08	0.146	0.005	0.54
127	rs12439433	0.966	0.02	0.034	0.000	0.51
128	rs735480	0.586	0.23	0.414	0.039	0.65
129	rs1426654	0.751	0.14	0.249	0.012	0.57
130	rs2899826	0.593	0.22	0.407	0.044	0.66
131	rs8035124	0.414	0.36	0.586	0.197	1.02
132	rs4984913	0.419	0.33	0.581	0.111	0.83
133	rs4781011	0.429	0.32	0.571	0.105	0.81
134	rs818386	0.798	0.10	0.202	0.010	0.56
135	rs2966849	0.758	0.12	0.242	0.015	0.58
136	rs459920	0.363	0.37	0.637	0.166	0.95
137	rs1879488	0.841	0.08	0.159	0.006	0.55
138	rs4411548	0.613	0.21	0.387	0.037	0.65
139	rs2593595	0.426	0.34	0.574	0.155	0.93
140	rs17642714	0.564	0.24	0.436	0.056	0.69
141	rs4471745	0.958	0.02	0.042	0.000	0.51
142	rs2033111	0.488	0.29	0.512	0.083	0.76
143	rs11652805	0.366	0.36	0.634	0.123	0.85
144	rs10512572	0.685	0.16	0.315	0.026	0.61
145	rs2125345	0.363	0.37	0.637	0.158	0.93
146	rs4798812	0.367	0.37	0.633	0.171	0.96
147	rs2042762	0.979	0.01	0.021	0.000	0.51
148	rs7226659	0.672	0.18	0.328	0.024	0.61

SNP#	SNP ID	PM	PIC	PD	PE	TPI
149	rs7238445	0.396	0.35	0.604	0.135	0.88
150	rs881728	0.539	0.25	0.461	0.061	0.71
151	rs3916235	0.612	0.21	0.388	0.038	0.65
152	rs4891825	0.572	0.23	0.428	0.050	0.68
153	rs874299	0.423	0.33	0.577	0.112	0.83
154	rs7251928	0.375	0.35	0.625	0.094	0.79
155	rs8113143	0.400	0.36	0.600	0.183	0.99
156	rs3745099	0.405	0.34	0.595	0.124	0.86
157	rs2532060	0.387	0.36	0.613	0.153	0.92
158	rs6104567	0.496	0.28	0.504	0.081	0.76
159	rs3907047	0.916	0.04	0.084	0.002	0.52
160	rs310644	0.616	0.21	0.384	0.035	0.64
161	rs2835370	0.591	0.22	0.409	0.052	0.69
162	rs1296819	0.431	0.33	0.569	0.147	0.91
163	rs4821004	0.371	0.37	0.629	0.165	0.95
164	rs2024566	0.425	0.34	0.575	0.146	0.91
165	rs5768007	0.629	0.20	0.371	0.034	0.64

9.14 Appendix 14: Admixture Prediction and Population likelihood values (population samples).

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)						Population Likelihoods			
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT 1	M	Umm Salal Municipality					20		80	Palestinian	Asia	1.27E-44
QAT 2	M	Al Khor and Al Thakira Municipality				10			90	Kuwaiti	Asia	1.10E-47
QAT 3	M	Al Sheehaniya Municipality							100	Jews, Yemenite	Asia	9.10E-43
QAT 4	M	Al Sheehaniya Municipality							100	Palestinian	Asia	7.73E-44
QAT 5	M	Umm Salal Municipality	5				40		55	Keralite	Asia	1.93E-43
QAT 6	M	Al Rayyan Municipality					15		85	Palestinian	Asia	6.82E-46
QAT 7	M	Doha Municipality		5			60		35	Pashtun	Asia	2.71E-44
QAT 8	M	Umm Salal Municipality		10	5	10			75	Negroid Makrani	Asia	1.85E-49
QAT 9	M	Al Rayyan Municipality	40	15					45	Kuwaiti	Asia	2.26E-48
QAT 10	M	Al Rayyan Municipality	20						80	Druze	Asia	4.02E-39
QAT 11	M	Umm Salal Municipality							100	Jews, Yemenite	Asia	3.79E-42
QAT 12	M	Al Wakra Municipality				10	40		50	Mohanna	Asia	1.08E-49
QAT 13	M	Al Khor and Al Thakira Municipality	5						95	Kuwaiti	Asia	3.49E-43
QAT 14	M	Al Rayyan Municipality							100	Kuwaiti	Asia	1.64E-46
QAT 15	M	Al Rayyan Municipality	35						65	Druze	Asia	5.74E-38
QAT 16	M	Al Khor and Al Thakira Municipality				10			90	Kuwaiti	Asia	1.28E-49
QAT 17	M	Doha Municipality	10				5		85	Druze	Asia	2.15E-44
QAT 18	M	Al Rayyan Municipality							100	Palestinian	Asia	3.40E-39
QAT 19	M	Al Rayyan Municipality							100	Jews, Yemenite	Asia	1.87E-45
QAT 20	M	Al Wakra Municipality							100	Jews, Yemenite	Asia	5.48E-42
QAT 21	M	Al Rayyan Municipality				20	10		70	Palestinian	Asia	2.67E-53
QAT 22	M	Al Wakra Municipality							100	Druze	Asia	1.32E-44
QAT 23	M	Al Rayyan Municipality							100	Palestinian	Asia	6.64E-45
QAT 24	M	Doha Municipality	10	10		5			75	Kuwaiti	Asia	2.02E-49
QAT 25	M	Umm Salal Municipality							100	Palestinian	Asia	3.93E-45
QAT 26	M	Al Rayyan Municipality					5		95	Palestinian	Asia	2.97E-43
QAT 27	M	Doha Municipality							100	Palestinian	Asia	2.96E-44
QAT 28	M	Al Rayyan Municipality	15			5	20		60	Kuwaiti	Asia	3.57E-47
QAT 29	M	Al Rayyan Municipality		5		10			85	Kuwaiti	Asia	1.07E-44
QAT 30	M	Al Rayyan Municipality	10						90	Kuwaiti	Asia	7.21E-44
QAT 31	M	Al Rayyan Municipality	5						95	Palestinian	Asia	5.05E-40
QAT 32	M	Al Wakra Municipality							100	Jews, Yemenite	Asia	6.67E-38

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)						Population Likelihoods			
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT 33	M	Al Rayyan Municipality		5					95	Kuwaiti	Asia	4.95E-44
QAT 34	M	Al Rayyan Municipality						5	95	Kuwaiti	Asia	2.53E-45
QAT 35	M	Al Rayyan Municipality		5					95	Palestinian	Asia	8.60E-44
QAT 36	M	Al Rayyan Municipality	10					10	80	Palestinian	Asia	9.28E-56
QAT 37	M	Al Shamal Municipality		15		5			80	Palestinian	Asia	3.50E-53
QAT 38	M	Al Rayyan Municipality	5						95	Jews, Yemenite	Asia	1.97E-38
QAT 39	M	Umm Salal Municipality					20		80	Kuwaiti	Asia	1.60E-44
QAT 40	M	Doha Municipality				5	65		35	Pashtun	Asia	5.40E-50
QAT 41	M	Al Sheehaniya Municipality							100	Palestinian	Asia	3.58E-46
QAT 42	M	Al Rayyan Municipality			5				95	Jews, Yemenite	Asia	1.49E-45
QAT 43	M	Umm Salal Municipality		10					90	Palestinian	Asia	5.79E-47
QAT 44	M	Al Sheehaniya Municipality	5						95	Jews, Yemenite	Asia	2.99E-43
QAT 45	M	Al Rayyan Municipality				5	45		45	Kuwaiti	Asia	1.95E-56
QAT 46	M	Umm Salal Municipality	35	5					60	Druze	Asia	4.53E-39
QAT 47	M	Al Rayyan Municipality	5				5		90	Druze	Asia	1.48E-41
QAT 48	M	Umm Salal Municipality				5			95	Palestinian	Asia	2.48E-47
QAT 49	M	Al Khor and Al Thakira Municipality							100	Palestinian	Asia	5.88E-44
QAT 50	M	Al Rayyan Municipality				25	30		45	Negroid Makrani	Asia	1.23E-49
QAR 51	M	Doha Municipality							100	Jews, Yemenite	Asia	6.84E-40
QAT 52	M	Al Rayyan Municipality					10		90	Kuwaiti	Asia	8.72E-42
QAT 53	M	Al Rayyan Municipality	25						75	Sardinian	Europe	9.12E-39
QAT 54	M	Al Rayyan Municipality							100	Jews, Yemenite	Asia	1.23E-42
QAT 55	M	Umm Salal Municipality							100	Druze	Asia	6.50E-42
QAT 56	M	Al Rayyan Municipality	5						95	Palestinian	Asia	7.04E-42
QAT 57	M	Al Shamal Municipality	40				10		50	Kuwaiti	Asia	1.15E-45
QAT 58	M	Al Rayyan Municipality				30	45		25	Negroid Makrani	Asia	8.38E-53
QAT 59	M	Doha Municipality	15			5	15		65	Druze	Asia	1.27E-44
QAT 60	M	Al Rayyan Municipality				35			65	Jews, Ethiopian	Africa	1.53E-51
QAT61	M	Doha Municipality	60	5			30		5	Adygei	Europe	1.88E-42
QAT62	M	Al Rayyan Municipality					20		80	Jews, Yemenite	Asia	2.73E-43
QAT63	M	Umm Salal Municipality	5						95	Jews, Yemenite	Asia	6.75E-41
QAT64	M	Umm Salal Municipality	15						85	Sardinian	Europe	3.25E-44
QAT65	M	Umm Salal Municipality	15				5		80	Negroid Makrani	Asia	3.96E-45
QAT66	M	Al Rayyan Municipality							100	Palestinian	Asia	1.41E-43
QAT67	M	Al Rayyan Municipality					15		85	Kuwaiti	Asia	5.10E-40
QAT68	M	Umm Salal Municipality					15		85	Kuwaiti	Asia	9.60E-45

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)						Population Likelihoods			
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT69	M	Al Rayyan Municipality		10					90	Palestinian	Asia	2.73E-44
QAT70	M	Al Rayyan Municipality	20						80	Jews, Yemenite	Asia	9.38E-46
QAT71	M	Al Rayyan Municipality	20						80	Jews, Yemenite	Asia	4.99E-43
QAT72	M	Al Sheehaniya Municipality	15			5			80	Kuwaiti	Asia	7.10E-47
QAT73	M	Doha Municipality					10		90	Palestinian	Asia	1.77E-41
QAT74	M	Umm Salal Municipality	10						90	Druze	Asia	5.70E-44
QAT75	M	Al Rayyan Municipality							100	Jews, Yemenite	Asia	9.97E-47
QAT76	M	Doha Municipality		5		10			85	Kuwaiti	Asia	5.38E-49
QAT77	M	Umm Salal Municipality			5				95	Kuwaiti	Asia	3.73E-39
QAT78	M	Doha Municipality			5		5	5	85	Kuwaiti	Asia	5.19E-45
QAT79	M	Doha Municipality				5			95	Kuwaiti	Asia	8.13E-47
QAT80	M	Al Wakra Municipality	45						55	Greeks	Europe	8.09E-44
QAT81	M	Al Shamal Municipality		5					95	Kuwaiti	Asia	8.55E-47
QAT82	M	Al Rayyan Municipality	10				5		85	Druze	Asia	3.72E-43
QAT83	M	Umm Salal Municipality	5						95	Druze	Asia	3.22E-39
QAT84	M	Al Rayyan Municipality	5						95	Jews, Yemenite	Asia	2.14E-35
QAT85	M	Al Rayyan Municipality	40			5			55	Kuwaiti	Asia	9.87E-46
QAT86	M	Al Rayyan Municipality							100	Palestinian	Asia	4.70E-44
QAT87	M	Al Wakra Municipality							100	Kuwaiti	Asia	1.62E-42
QAT88	M	Umm Salal Municipality	10						90	Sardinian	Europe	1.77E-41
QAT89	M	Al Rayyan Municipality		20		10			70	Negroid Makrani	Asia	1.30E-48
QAT90	M	Al Rayyan Municipality							95	Palestinian	Asia	1.51E-42
QAT91	M	Doha Municipality					35		65	Pashtun	Asia	6.82E-48
QAT92	M	Doha Municipality					35		65	Kuwaiti	Asia	4.57E-48
QAT93	M	Al Rayyan Municipality	10	5					85	Palestinian	Asia	2.40E-49
QAT94	M	Al Rayyan Municipality				10			90	Kuwaiti	Asia	3.04E-50
QAT95	M	Al Rayyan Municipality		5			20		75	Kuwaiti	Asia	1.47E-47
QAT96	M	Al Rayyan Municipality	30		5				65	Jews, Sephardic	Europe	4.48E-40
QAT97	M	Al Rayyan Municipality					15		85	Jews, Yemenite	Asia	2.27E-46
QAT98	M	Al Sheehaniya Municipality	15						85	Jews, Sephardic	Europe	5.95E-43
QAT99	M	Doha Municipality	20				20		60	Palestinian	Asia	6.03E-42
QAT100	M	Doha Municipality	25				30		45	Jews, Sephardic	Europe	5.11E-45
QAT101	M	Al Rayyan Municipality		5					95	Palestinian	Asia	8.57E-42
QAT102	M	Doha Municipality						5	95	Palestinian	Asia	1.06E-51
QAT103	M	Al Rayyan Municipality					15		85	Palestinian	Asia	5.19E-46
QAT104	M	Al Khor and Al Thakira Municipality				5			95	Palestinian	Asia	2.41E-47

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)						Population Likelihoods			
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT105	M	Umm Salal Municipality	5						95	Jews, Yemenite	Asia	1.10E-38
QAT106	M	Al Khor and Al Thakira Municipality	5	5					90	Kuwaiti	Asia	3.42E-44
QAT107	M	Al Rayyan Municipality	20				10		70	Palestinian	Asia	5.37E-43
QAT108	M	Al Rayyan Municipality							100	Kuwaiti	Asia	4.13E-43
QAT109	M	Al Rayyan Municipality							100	Druze	Asia	2.23E-42
QAT110	M	Doha Municipality	15			10			75	Palestinian	Asia	5.15E-49
QAT111	M	Doha Municipality							100	Palestinian	Asia	6.13E-47
QAT112	M	Al Rayyan Municipality							100	Jews, Yemenite	Asia	4.91E-35
QAT113	M	Al Rayyan Municipality				5			95	Palestinian	Asia	3.76E-48
QAT114	M	Al Rayyan Municipality							100	Palestinian	Asia	1.18E-43
QAT115	M	Al Rayyan Municipality					5		95	Kuwaiti	Asia	2.94E-44
QAT116	M	Al Rayyan Municipality	20			5			75	Palestinian	Asia	1.24E-46
QAT117	M	Al Rayyan Municipality	10						90	Palestinian	Asia	2.56E-46
QAT118	M	Doha Municipality	40			20			40	Negroid Makrani	Asia	1.00E-51
QAT119	M	Umm Salal Municipality							100	Palestinian	Asia	4.56E-38
QAT120	M	Umm Salal Municipality	15					5	80	Kuwaiti	Asia	2.47E-48
QAT121	M	Al Khor and Al Thakira Municipality	5				10		85	Druze	Asia	1.65E-48
QAT122	M	Doha Municipality	15				70		15	Pashtun	Asia	2.33E-47
QAT123	M	Al Sheehaniya Municipality						10	90	Palestinian	Asia	4.74E-47
QAT124	M	Doha Municipality	45						55	Palestinian	Asia	1.72E-45
QAT125	M	Al Rayyan Municipality	25						65	Kuwaiti	Asia	1.51E-33
QAT126	M	Al Khor and Al Thakira Municipality		5			5		90	Kuwaiti	Asia	4.92E-35
QAT127	M	Al Rayyan Municipality						10	90	Kuwaiti	Asia	4.26E-35
QAT128	M	Doha Municipality	5					5	90	Jews, Yemenite	Asia	5.34E-31
QAT129	M	Al Daayen Municipality	15				10		75	Druze	Asia	5.70E-41
QAT130	M	Umm Salal Municipality	10			5			85	Kuwaiti	Asia	7.36E-51
QAT131	M	Al Rayyan Municipality				20			80	Kuwaiti	Asia	9.26E-42
QAT132	M	Umm Salal Municipality				60			40	African Americans	Africa	1.11E-51
QAT133	M	Al Rayyan Municipality		15	15	15			55	Jews, Ethiopian	Africa	2.84E-07
QAT134	M	Al Rayyan Municipality				10		5	85	Negroid Makrani	Asia	5.37E-42
QAT135	M	Umm Salal Municipality	10						90	Jews, Yemenite	Asia	7.66E-38
QAT136	M	Al Daayen Municipality						10	90	Kuwaiti	Asia	4.23E-38
QAT137	M	Doha Municipality						5	95	Palestinian	Asia	1.31E-47
QAT138	M	Al Rayyan Municipality	5						95	Samaritans	Europe	3.13E-41
QAT139	M	Al Rayyan Municipality					25		75	Jews, Sephardic	Europe	2.24E-45
QAT140	M	Umm Salal Municipality	40						60	Druze	Asia	4.97E-43

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)						Population Likelihoods			
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT141	M	Al Rayyan Municipality		5		40	5		50	Jews, Ethiopian	Africa	5.83E-50
QAT142	M	Doha Municipality					20		80	Palestinian	Asia	2.86E-42
QAT143	M	Al Wakra Municipality							100	Palestinian	Asia	1.85E-40
QAT144	M	Doha Municipality	15					10	75	Kuwaiti	Asia	1.89E-46
QAT145	M	Al Rayyan Municipality					20		80	Palestinian	Asia	7.08E-45
QAT146	M	Al Khor and Al Thakira Municipality	5		5				90	Jews, Yemenite	Asia	8.88E-45
QAT147	M	Al Shamal Municipality						5	95	Palestinian	Asia	1.08E-50
QAT148	M	Al Rayyan Municipality				15		20	65	Palestinian	Asia	1.04E-34
QAT149	M	Al Rayyan Municipality					15		85	Palestinian	Asia	6.78E-41
QAT150	M	Doha Municipality	25				45	5	25	Pashtun	Asia	5.24E-46
QAT151	M	Al Sheehaniya Municipality	30			25		15	35	Jews, Ethiopian	Africa	1.76E-35
QAT152	M	Al Rayyan Municipality		10					90	Palestinian	Asia	8.46E-42
QAT153	M	Doha Municipality	25				10		65	Jews, Sephardic	Europe	1.16E-42
QAT154	M	Umm Salal Municipality	15			15			70	Kuwaiti	Asia	7.46E-49
QAT155	M	Al Rayyan Municipality				15	25	5	55	Somali	Africa	2.00E-32
QAT156	M	Al Rayyan Municipality	15			5		15	65	Palestinian	Asia	8.57E-27
QAT157	M	Doha Municipality	10			15		15	65	Negroid Makrani	Asia	1.36E-32
QAT158	M	Al Khor and Al Thakira Municipality					30		70	Druze	Asia	2.06E-38
QAT159	M	Al Rayyan Municipality					15		85	Palestinian	Asia	3.32E-37
QAT160	M	Al Wakra Municipality	35				50		15	Pashtun	Asia	1.05E-48
QAT161	M	Al Sheehaniya Municipality					5		95	Palestinian	Asia	3.06E-45
QAT162	M	Al Wakra Municipality	40				25		35	Greeks	Europe	1.18E-41
QAT163	M	Umm Salal Municipality	35						65	Jews, Sephardic	Europe	3.56E-40
QAT164	M	Doha Municipality	25	5			5		65	Kuwaiti	Asia	1.97E-46
QAT165	M	Umm Salal Municipality	5	5		10			80	Kuwaiti	Asia	5.77E-49
QAT166	M	Al Wakra Municipality	25	5					70	Palestinian	Asia	5.48E-47
QAT167	M	Al Khor and Al Thakira Municipality							100	Palestinian	Asia	2.05E-47
QAT168	M	Doha Municipality		5			5		90	Palestinian	Asia	5.10E-40
QAT169	M	Doha Municipality					25		75	Pashtun	Asia	4.98E-49
QAT170	M	Umm Salal Municipality							100	Kuwaiti	Asia	7.87E-41
QAT171	M	Doha Municipality				5	20	5	70	Negroid Makrani	Asia	2.00E-42
QAT172	M	Al Rayyan Municipality	5			30		10	55	Somali	Africa	1.46E-17
QAT173	M	Al Rayyan Municipality	5					5	90	Palestinian	Asia	8.91E-49
QAT174	M	Al Rayyan Municipality				5			95	Kuwaiti	Asia	7.14E-41
QAT175	M	Doha Municipality	5				50		45	Pashtun	Asia	1.65E-38
QAT176	M	Al Wakra Municipality	5				5		90	Druze	Asia	1.10E-42

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)							Population Likelihoods		
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT177	F	Doha Municipality	10				10	5	75	Kuwaiti	Asia	2.15E-49
QAT178	F	Al Rayyan Municipality	10						90	Kuwaiti	Asia	2.70E-44
QAT179	F	Al Rayyan Municipality		5		5			90	Palestinian	Asia	8.42E-54
QAT180	F	Al Daayen Municipality				35			65	Jews, Ethiopian	Africa	2.50E-46
QAT181	M	Al Wakra Municipality	5				25	5	65	Pashtun	Asia	5.21E-48
QAT182	F	Doha Municipality		5	5	65	25			African Americans	Africa	1.45E-49
QAT183	F	Al Rayyan Municipality							100	Palestinian	Asia	1.50E-44
QAT184	F	Doha Municipality				35			65	Jews, Ethiopian	Africa	2.82E-48
QAT185	M	Doha Municipality	10		5				85	Jews, Yemenite	Asia	4.67E-48
QAT186	M	Al Wakra Municipality					25		75	Palestinian	Asia	2.07E-44
QAT187	F	Al Rayyan Municipality	50						50	Adygei	Europe	1.15E-41
QAT188	F	Al Rayyan Municipality					5		95	Palestinian	Asia	6.73E-41
QAT189	F	Al Rayyan Municipality				5	5		90	Jews, Yemenite	Asia	3.13E-47
QAT190	M	Al Rayyan Municipality							100	Kuwaiti	Asia	1.51E-43
QAT191	M	Doha Municipality	10						90	Palestinian	Asia	5.04E-45
QAT192	F	Al Rayyan Municipality	10						90	Palestinian	Asia	1.52E-42
QAT193	F	Al Rayyan Municipality							100	Palestinian	Asia	1.24E-44
QAT194	F	Al Rayyan Municipality						10	90	Jews, Yemenite	Asia	1.03E-41
QAT195	F	Al Rayyan Municipality				75			25	African Americans	Africa	2.53E-43
QAT196	F	Al Rayyan Municipality				10	15		75	Palestinian	asia	1.79E-50
QAT197	M	Al Rayyan Municipality		10		5			85	Jews, Yemenite	Asia	7.57E-49
QAT198	M	Al Rayyan Municipality							100	Jews, Yemenite	Asia	7.84E-41
QAT199	M	Umm Salal Municipality				5	45		50	Kuwaiti	Asia	2.64E-50
QAT200	M	Doha Municipality	30		10			5	55	Pashtun	Asia	5.28E-50
QAT201	M	Doha Municipality							100	Jews, Yemenite	Asia	2.90E-45
QAT202	M	Al Rayyan Municipality	15			10	45		30	Kuwaiti	Asia	4.84E-45
QAT203	M	Al Rayyan Municipality	30			15			55	Kuwaiti	Asia	3.34E-51
QAT204	M	Al Wakra Municipality				30	25		45	Somali	Africa	5.62E-39
QAT205	F	Umm Salal Municipality		10	5	50			35	African Americans	Africa	1.04E-55
QAT206	F	Al Wakra Municipality	20				20		60	Palestinian	Asia	4.32E-45
QAT207	M	Al Rayyan Municipality							100	Kuwaiti	Asia	5.17E-44
QAT208	F	Al Rayyan Municipality	15			5			80	Jews, Yemenite	Asia	3.37E-48
QAT209	F	Doha Municipality	15						85	Jews, Yemenite	Asia	2.49E-49
QAT210	F	Umm Salal Municipality	30			15	20		35	Negroid Makrani	Asia	6.56E-55
QAT211	M	Doha Municipality	10						90	Palestinian	Asia	1.88E-42
QAT212	F	Al Rayyan Municipality		10		15	10		65	Somali	Africa	2.84E-54

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)						Population Likelihoods			
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT213	F	Al Rayyan Municipality	20	5					75	Druze	Asia	1.38E-43
QAT214	M	Al Rayyan Municipality	15	5			65		15	Pashtun	Asia	1.26E-46
QAT215	M	Al Rayyan Municipality							100	Kuwaiti	Asia	4.54E-40
QAT216	M	Doha Municipality		5					95	Kuwaiti	Asia	4.13E-43
QAT217	M	Al Rayyan Municipality	50	10			25		15	Jews, Ashkenazi	Europe	3.54E-47
QAT218	M	Al Rayyan Municipality							100	Palestinian	Asia	4.57E-40
QAT219	M	Doha Municipality	25			10	30		35	Palestinian	Asia	7.05E-52
QAT220	M	Al Wakra Municipality					90		10	Sandawe	Africa	7.20E-38
QAT221	M	Al Daayen Municipality			5				95	Kuwaiti	Asia	4.98E-44
QAT222	F	Al Rayyan Municipality	35			15	10		40	Palestinian	Asia	2.44E-48
QAT223	M	Doha Municipality				15	65		20	Pashtun	Asia	3.30E-51
QAT224	M	Doha Municipality							100	Kuwaiti	Asia	8.56E-46
QAT225	F	Al Daayen Municipality				5	45		55	Pashtun	Asia	5.48E-54
QAT226	F	Doha Municipality			10				90	Palestinian	Asia	2.34E-47
QAT227	M	Al Rayyan Municipality							100	Palestinian	Asia	1.73E-40
QAT228	F	Doha Municipality	25	10					65	Jews, Sephardic	Europe	2.12E-45
QAT229	M	Al Wakra Municipality	10				25		65	Palestinian	Asia	2.01E-43
QAT230	M	Al Rayyan Municipality					5		95	Druze	Asia	1.16E-41
QAT231	M	Al Rayyan Municipality							100	Palestinian	Asia	8.90E-37
QAT232	M	Al Wakra Municipality	5			20		5	70	Kuwaiti	Asia	4.25E-48
QAT233	F	Doha Municipality	10			20			70	Palestinian	Asia	1.46E-51
QAT234	M	Al Rayyan Municipality		5					95	Palestinian	Asia	1.28E-42
QAT235	M	Doha Municipality	25						75	Jews, Yemenite	Asia	6.39E-37
QAT236	M	Al Sheehaniya Municipality		15		5	45		40	Keralite	Asia	1.08E-49
QAT237	F	Doha Municipality				5	20		75	Negroid Makrani	Asia	9.70E-44
QAT238	F	Al Rayyan Municipality			5	15			80	Palestinian	Asia	1.29E-50
QAT239	F	Al Rayyan Municipality							100	Jews, Yemenite	Asia	2.35E-43
QAT240	F	Al Rayyan Municipality	5			10			85	Kuwaiti	Asia	7.33E-47
QAT241	F	Al Rayyan Municipality				65			35	African Americans	Africa	1.22E-45
QAT242	F	Al Rayyan Municipality				5			95	Palestinian	Asia	2.93E-43
QAT243	F	Al Rayyan Municipality							100	Jews, Yemenite	Asia	6.57E-39
QAT244	F	Al Rayyan Municipality	15			10			75	Palestinian	Asia	1.85E-49
QAT245	M	Al Shamal Municipality	20						80	Jews, Yemenite	Asia	2.13E-40
QAT246	F	Doha Municipality					40		60	Palestinian	Asia	1.77E-45
QAT247	M	Al Rayyan Municipality	10						90	Druze	Asia	1.44E-41
QAT248	F	Umm Salal Municipality	20				5		75	Druze	Asia	1.02E-41

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)						Population Likelihoods			
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT249	M	Al Daayen Municipality				10			90	Palestinian	Asia	1.79E-47
QAT250	F	Doha Municipality	20					10	70	Palestinian	Asia	1.45E-39
QAT251	F	Doha Municipality	30	15				40	15	Pashtun	Asia	1.69E-45
QAT252	F	Doha Municipality	20					55	25	Pashtun	Asia	3.59E-48
QAT253	M	Al Wakra Municipality	5					15	80	Kuwaiti	Asia	2.37E-43
QAT254	M	Al Rayyan Municipality				80			15	African Americans	Africa	4.80E-43
QAT255	F	Doha Municipality				10			90	Palestinian	Asia	2.52E-48
QAT256	M	Al Rayyan Municipality				15	30		55	Kuwaiti	Asia	4.17E-50
QAT257	M	Doha Municipality	15				30		55	Palestinian	Asia	2.92E-47
QAT258	M	Al Rayyan Municipality	10				60		30	Pashtun	Asia	1.53E-45
QAT259	F	Al Rayyan Municipality							100	Kuwaiti	Asia	1.46E-46
QAT260	F	Al Rayyan Municipality							100	Sardinian	Europe	6.71E-40
QAT261	M	Al Rayyan Municipality		5					95	Palestinian	Asia	7.89E-45
QAT262	M	Doha Municipality	35						65	Druze	Asia	2.73E-41
QAT263	M	Doha Municipality					5	5	90	Kuwaiti	Asia	8.71E-47
QAT264	M	Al Rayyan Municipality	15						85	Jews, Yemenite	Asia	4.83E-39
QAT265	M	Al Rayyan Municipality				5			95	Jews, Yemenite	Asia	4.16E-50
QAT266	M	Umm Salal Municipality					30		70	Kuwaiti	Asia	1.47E-41
QAT267	M	Al Wakra Municipality	5				35		65	Kuwaiti	Asia	4.25E-46
QAT268	M	Al Rayyan Municipality					10		90	Kuwaiti	Asia	1.44E-48
QAT269	M	Doha Municipality	50			10	5		35	Kuwaiti	Asia	1.83E-48
QAT270	M	Al Rayyan Municipality					20		80	Jews, Yemenite	Asia	8.07E-47
QAT271	F	Al Rayyan Municipality	5			15	5		75	Kuwaiti	Asia	1.37E-53
QAT272	M	Al Rayyan Municipality	10						90	Jews, Yemenite	Asia	3.25E-45
QAT273	M	Umm Salal Municipality			5	5			90	Kuwaiti	Asia	3.28E-46
QAT274	M	Doha Municipality			15				85	Druze	Asia	2.15E-56
QAT275	F	Doha Municipality	30			10	5		55	Kuwaiti	Asia	8.16E-49
QAT276	M	Al Rayyan Municipality		5					95	Druze	Asia	1.10E-43
QAT277	M	Doha Municipality	20				80			Pashtun	Asia	3.17E-45
QAT278	M	Al Rayyan Municipality					5		95	Kuwaiti	Asia	3.84E-44
QAT279	M	Doha Municipality							100	Palestinian	Asia	2.32E-43
QAT280	M	Al Rayyan Municipality				5			95	Palestinian	Asia	3.87E-43
QAT281	M	Umm Salal Municipality	30						70	Palestinian	Asia	8.83E-44
QAT282	M	Al Rayyan Municipality							100	Jews, Yemenite	Asia	6.76E-43
QAT283	F	Doha Municipality	15						85	Palestinian	Asia	6.04E-44
QAT284	F	Doha Municipality							100	Jews, Yemenite	Asia	1.01E-37

Sample name	Gender	Municipality name	Admixture Prediction results (population name%)							Population Likelihoods		
			Europe	Oceania	East Asia	Africa	South Asia	America	Southwest Asia	Population Name	Geo Region	Likelihood
QAT285	M	Doha Municipality	15				10		75	Druze	Asia	2.33E-43
QAT286	M	Al Daayen Municipality	10	5		25			60	Negroid Makrani	Asia	1.23E-48
QAT287	M	Doha Municipality	10						90	Palestinian	Asia	1.71E-41
QAT288	M	Al Rayyan Municipality	10			15	70		5	Negroid Makrani	Asia	2.63E-48
QAT289	M	Doha Municipality				5		5	90	Palestinian	Asia	4.03E-48
QAT290	M	Al Rayyan Municipality							100	Palestinian	Asia	2.03E-43
QAT291	M	Al Rayyan Municipality							100	Jews, Yemenite	Asia	3.08E-46
QAT292	M	Doha Municipality	15			10	15		60	Palestinian	Asia	2.13E-54
QAT293	M	Al Khor and Al Thakira Municipality							100	Jews, Yemenite	Asia	3.90E-38
QAT294	M	Doha Municipality		5					95	Palestinian	Asia	2.03E-44
QAT295	M	Doha Municipality	15			10			75	Negroid Makrani	Asia	6.77E-49
QAT296	M	Doha Municipality		5		5	30		60	Negroid Makrani	Asia	3.51E-46
QAT297	M	Doha Municipality							100	Palestinian	Asia	1.24E-44
QAT298	M	Al Daayen Municipality	10						90	Druze	Asia	3.69E-43
QAT299*	M	Al Daayen Municipality			5				95	Kuwaiti	Asia	4.73E-49
QAT300	M	Umm Salal Municipality							100	Jews, Yemenite	Asia	8.47E-40

*Sample QAT299 was re-analyzed and the result added to the table.

9.15 Appendix 15: Data showing the combined results of the 105 Qatari samples from Precision ID Ancestry and Identity panel experiments.

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S16-BC16-P1	America	3.38E-37	R2	Palestinian	In data
	Europe	1.22E-37			
	South Asia	3.11E-41			
	East Asia	3.41E-42			
	Africa	2.07E-49			
S17-BC17-P1	America	1.47E-40	J	Keralite	In data
	South Asia	3.95E-41			
	Europe	3.05E-41			
	East Asia	1.21E-41			
	Africa	4.09E-46			
S18-BC18-P1	America	7.96E-38	J	Jews, Yemenite	In data
	South Asia	6.69E-38			
	Europe	3.26E-38			
	East Asia	2.43E-39			
	Africa	6.56E-47			
S19-BC19-P1	America	6.69E-39	J	Palestinian	In data
	South Asia	2.08E-40			
	Europe	1.32E-42			
	East Asia	1.55E-43			
	Africa	9.49E-44			
S20-BC20-P1	America	2.62E-35	J	Palestinian	In data
	South Asia	2.87E-36			
	Africa	1.20E-36			
	Europe	7.87E-37			
	East Asia	3.53E-37			

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S21-BC21-P1	America Europe South Asia East Asia Africa	1.72E-37 3.96E-38 1.51E-38 5.05E-39 6.80E-42	R2	Druze	In data
S22-BC22-P1	America Europe South Asia East Asia Africa	4.13E-38 3.25E-38 3.50E-41 3.02E-42 9.07E-43	J	Jews, Yemenite	In data
S23-BC23-P1	America Europe South Asia East Asia Africa	9.09E-41 5.71E-41 3.88E-42 8.50E-44 1.65E-51	J	Druze	In data
S24-BC24-P1	East Asia South Asia America Europe Africa	4.77E-35 2.05E-35 1.05E-35 1.20E-36 1.14E-38	R1a1	Negroid Makrani	In data
S25-BC25-P1	America Africa South Asia Europe East Asia	1.69E-36 1.61E-37 1.38E-37 5.17E-38 1.46E-41	J	Kuwaiti	In data
S26-BC26-P1	Europe America South Asia East Asia Africa	1.90E-38 5.89E-39 9.95E-42 4.05E-44 1.12E-50	T	Sardinian	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S27-BC27-P1	America Europe South Asia East Asia Africa	2.01E-37 1.16E-37 1.09E-38 1.11E-40 1.24E-49	J	Kuwaiti	In data
S28-BC28-P1	South Asia America Europe Africa East Asia	6.64E-39 1.19E-39 9.97E-41 1.51E-41 1.02E-42	Female	African Americans	African Americans
S29-BC29-P1	Europe America South Asia East Asia Africa	4.18E-39 1.66E-39 5.17E-42 2.25E-42 1.73E-46	Female	Negroid Makrani	In data
S30-BC30-P1	Europe South Asia America East Asia Africa	1.43E-40 4.00E-41 1.97E-42 1.29E-45 5.56E-50	J	Kuwaiti	In data
S1-BC1-P2	America East Asia Europe South Asia Africa	8.80E-38 6.13E-38 6.12E-38 1.52E-39 3.47E-42	R1a1	Palestinian	In data
S2-BC2-P2	Europe America East Asia South Asia Africa	8.65E-39 4.07E-39 1.25E-42 2.78E-43 1.87E-44	J	Palestinian	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S3-BC3-P2	America Europe South Asia East Asia Africa	3.30E-38 2.05E-39 5.55E-41 6.20E-43 8.47E-45	J	Kuwaiti	In data
S4-BC4-P2	America Europe South Asia East Asia Africa	4.84E-39 4.53E-40 1.38E-40 5.17E-42 1.42E-47	J	Kuwaiti	In data
S5-BC5-P2	America Europe East Asia South Asia Africa	8.72E-38 1.30E-38 6.63E-40 3.84E-40 8.56E-47	J	Palestinian	In data
S6-BC6-P2	Europe America South Asia East Asia Africa	3.35E-38 8.87E-39 6.60E-41 4.55E-45 1.32E-46	J	Jews, Yemenite	In data
S7-BC7-P2	Europe America South Asia Africa East Asia	1.19E-38 4.06E-39 1.96E-40 3.19E-44 3.11E-44	J	Palestinian	In data
S8-BC8-P2	South Asia Europe America East Asia Africa	3.71E-36 7.65E-38 3.51E-38 2.65E-39 2.40E-40	T	Kuwaiti	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S9-BC9-P2	Europe America South Asia East Asia Africa	4.63E-36 6.17E-37 3.52E-38 4.44E-43 1.01E-44	J	Druze	In data
S10-BC10-P2	America Europe East Asia South Asia Africa	2.85E-39 1.34E-39 3.19E-40 2.68E-42 1.78E-46	J	Kuwaiti	In data
S11-BC11-P2	South Asia America East Asia Europe Africa	1.90E-37 6.67E-40 5.79E-40 1.70E-40 1.87E-46	J	Jews, Yemenite	In data
S12-BC12-P2	Europe America South Asia East Asia Africa	1.32E-37 1.30E-39 1.29E-39 3.28E-43 1.12E-45	Q	Palestinian	In data
S13-BC13-P2	Europe America East Asia South Asia Africa	2.08E-38 2.07E-39 1.66E-40 1.58E-41 1.55E-46	J	Jews, Yemenite	Jews, Yemenite
S14-BC14-P2	America Europe East Asia South Asia Africa	2.56E-37 1.62E-37 8.57E-38 6.78E-40 6.97E-45	J	Kuwaiti	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S15-BC15-P2	East Asia America South Asia Europe Africa	8.31E-37 3.21E-38 2.37E-38 1.85E-38 6.58E-48	J	Kuwaiti	In data
S16-BC16-P2	America Europe South Asia East Asia Africa	7.75E-37 2.97E-37 5.65E-38 1.04E-40 7.28E-43	J	Palestinian	In data
S17-BC17-P2	Europe South Asia America East Asia Africa	4.15E-40 6.88E-41 9.19E-42 1.49E-43 6.57E-46	Female	Kuwaiti	In data
S18-BC18-P2	Europe America South Asia East Asia Africa	1.32E-37 1.06E-37 7.30E-39 3.56E-44 4.63E-47	Female	Palestinian	In data
S19-BC19-P2	Europe America South Asia East Asia Africa	1.47E-37 4.79E-38 9.84E-40 9.84E-40 1.46E-47	Female	Palestinian	In data
S20-BC20-P2	Europe America South Asia East Asia Africa	1.71E-37 1.95E-38 1.57E-40 1.63E-41 2.94E-47	Female	Palestinian	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S21-BC21-P2	America Europe East Asia South Asia Africa	6.34E-36 1.83E-36 1.12E-39 2.84E-40 9.09E-43	Female	Jews, Yemenite	In data
S22-BC22-P2	Europe America South Asia East Asia Africa	1.05E-44 2.67E-46 2.18E-49 9.06E-50 5.91E-56	Female	Palestinian	Palestinian
S23-BC23-P2	Europe East Asia America South Asia Africa	1.59E-38 2.34E-39 1.94E-39 1.26E-39 1.62E-47	Female	Jews, Yemenite	In data
S24-BC24-P2	America South Asia Europe Africa East Asia	2.38E-38 1.33E-39 4.34E-40 2.71E-41 3.57E-43	Female	Palestinian	In data
S25-BC25-P2	Europe South Asia America East Asia Africa	3.47E-41 2.71E-41 7.90E-42 4.33E-42 3.22E-44	J	Kuwaiti	In data
S26-BC26-P2	America Europe East Asia South Asia Africa	1.79E-35 2.40E-36 6.58E-37 7.84E-38 6.52E-42	Female	Kuwaiti	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S27-BC27-P2	Europe America East Asia South Asia Africa	3.25E-40 1.67E-40 2.93E-42 1.79E-42 6.45E-49	Female	Kuwaiti	In data
S28-BC28-P2	Europe America South Asia East Asia Africa	6.68E-37 1.58E-38 7.54E-40 8.83E-43 3.95E-51	J	Druze	In data
S29-BC29-P2	Europe South Asia America East Asia Africa	4.20E-39 1.10E-39 3.15E-40 5.82E-44 2.29E-45	F	Kuwaiti	In data
S30-BC30-P2	America Europe South Asia East Asia Africa	3.90E-39 9.03E-40 1.33E-40 5.25E-42 1.7E-44	J	Pashtun	In data
S1-BC1-P3	South Asia East Asia Europe America Africa	1.31E-36 7.07E-37 2.65E-37 1.35E-37 1.23E-40	E	Druze	In data
S2-BC2-P3	Europe America South Asia East Asia Africa	2.27E-37 1.75E-37 8.25E-39 6.18E-43 5.5E-46	J	Adygei	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S3-BC3-P3	East Asia America South Asia Europe Africa	3.51E-39 8.58E-40 5.24E-40 2.23E-41 1.79E-48	J	Kuwaiti	In data
S4-BC4-P3	Europe America South Asia East Asia Africa	1.19E-34 2.89E-35 2.47E-36 1.63E-38 1.22E-39	B	Pashtun	In data
S5-BC5-P3	America Europe South Asia East Asia Africa	7.55E-37 3.66E-37 1.43E-37 1.52E-38 8.76E-45	E	Kuwaiti	In data
S6-BC6-P3	Europe South Asia America East Asia Africa	5.52E-36 8.48E-37 1.89E-37 2.76E-40 1.85E-44	G	Palestinian	In data
S7-BC7-P3	America Europe East Asia South Asia Africa	1.04E-37 5.97E-38 2.68E-39 1.19E-39 2.66E-43	R1a1	Pashtun	In data
S8-BC8-P3	Europe America South Asia East Asia Africa	5.82E-36 2.54E-37 6.69E-39 9.09E-40 1.81E-43	Female	Kuwaiti	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S9-BC9-P3	Africa East Asia South Asia America Europe	4.07E-35 1.23E-37 2.86E-38 1.73E-39 2.01E-41	Female	African Americans	In data
S10-BC10-P3	America Europe East Asia South Asia Africa	9.79E-36 7.43E-37 2.21E-37 5.62E-38 5.39E-39	Female	Jews, Ethiopian	In data
S11-BC11-P3	Europe South Asia East Asia America Africa	3.90E-43 5.25E-44 3.53E-45 7.78E-46 4.62E-49	J	Jews, Yemenite	In data
S12-BC12-P3	Europe America East Asia South Asia Africa	1.72E-41 7.47E-44 5.43E-46 3.19E-46 3.52E-52	J	Palestinian	Qatar
S13-BC13-P3	America Europe South Asia East Asia Africa	1.20E-34 1.46E-35 1.24E-36 3.78E-38 1.64E-41	Female	Negroid Makrani	In data
S14-BC14-P3	America Europe South Asia Africa East Asia	1.44E-37 1.51E-38 1.73E-39 1.70E-41 1.70E-41	Female	Palestinian	Qatar

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S15-BC15-P3	South Asia America Europe East Asia Africa	1.62E-38 7.51E-39 3.29E-39 6.02E-43 2.38E-43	Female	Kuwaiti	Kuwaiti
S16-BC16-P3	America Europe East Asia South Asia Africa	3.43E-38 1.59E-39 4.17E-40 2.32E-40 1.76E-45	L	Mohanna	In data
S17-BC17-P3	Europe America South Asia East Asia Africa	2.29E-37 1.24E-38 5.97E-40 1.48E-40 2.24E-47	R1a1	Jews, Yemenite	In data
S18-BC18-P3	Europe America South Asia East Asia Africa	9.45E-39 2.15E-39 9.71E-41 1.45E-44 1.38E-46	R1a1	Druze	In data
S19-BC19-P3	Europe America South Asia East Asia Africa	1.90E-36 8.5E-37 1.21E-38 9.21E-39 3.02E-45	R1a1	Jews, Yemenite	In data
S20-BC20-P3	Europe America South Asia East Asia Africa	2.26E-39 9.64E-40 5.75E-42 4.07E-42 1.09E-50	J	Greeks	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S21-BC21-P3	Europe America South Asia East Asia Africa	4.73E-39 2.41E-39 4.03E-42 7.02E-45 2.19E-48	J	Kuwaiti	In data
S22-BC22-P3	Europe America South Asia East Asia Africa	1.12E-34 6.05E-36 3.41E-37 8.01E-43 3.26E-47	R1a1	Palestinian	Qatar
S23-BC23-P3	Europe America South Asia East Asia Africa	1.03E-36 4.70E-37 1.05E-37 1.38E-40 5.84E-42	J	Pashtun	In data
S24-BC24-P3	South Asia America Europe East Asia Africa	7.71E-39 4.54E-39 1.58E-39 4.32E-40 2.08E-42	J	Palestinian	In data
S25-BC25-P3	Europe America South Asia East Asia Africa	1.64E-40 2.09E-42 1.27E-43 1.70E-45 1.39E-51	J	Kuwaiti	In data
S26-BC26-P3	Europe America South Asia East Asia Africa	6.13E-39 1.06E-39 3.70E-40 1.21E-42 1.67E-44	J	Kuwaiti	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S27-BC27-P3	America South Asia East Asia Europe Africa	1.92E-37 1.23E-37 1.20E-37 9.52E-38 4.54E-41	R1a1	Kuwaiti	In data
S28-BC28-P3	Europe America South Asia East Asia Africa	3.19E-41 1.43E-41 6.48E-42 1.94E-45 1.28E-50	J	Kuwaiti	In data
S29-BC29-P3	Europe South Asia America East Asia Africa	3.88E-39 6.57E-40 2.01E-40 1.35E-43 2.86E-46	J	Kuwaiti	In data
S30-BC30-P3	America Europe East Asia South Asia Africa	9.71E-39 4.61E-39 2.21E-39 4.27E-40 2.45E-43	J	Palestinian	In data
S1-BC1-P4	America Europe South Asia East Asia Africa	5.23E-40 9.05E-41 4.48E-43 2.16E-45 4.47E-50	R1a1	Palestinian	In data
S2-BC2-P4	Europe America South Asia East Asia Africa	1.03E-35 1.38E-36 2.56E-37 6.87E-38 3.09E-45	J	Kuwaiti	Kuwaiti

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S3-BC3-P4	Europe America South Asia East Asia Africa	8.60E-38 4.06E-38 3.24E-38 6.34E-41 7.68E-43	J	Druze	In data
S4-BC4-P4	Europe America South Asia East Asia Africa	3.51E-38 3.51E-38 5.27E-41 9.03E-44 1.63E-45	J	Kuwaiti	Qatar
S5-BC5-P4	America East Asia Europe South Asia Africa	1.45E-35 2.65E-36 1.91E-36 9.56E-37 6.24E-39	J	Jews, Yemenite	In data
S6-BC6-P4	Europe South Asia America Africa East Asia	1.85E-37 4.35E-38 1.62E-39 9.38E-41 1.04E-42	J	Druze	Qatar
S7-BC7-P4	America South Asia Europe East Asia Africa	2.68E-36 8.91E-37 5.59E-37 2.61E-39 1.22E-43	J	Jews, Yemenite	In data
S8-BC8-P4	Europe America South Asia East Asia Africa	2.82E-38 8.82E-39 3.79E-41 1.11E-41 3.35E-49	J	Jews, Yemenite	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S9-BC9-P4	America Europe South Asia East Asia Africa	1.62E-36 1.13E-37 3.41E-38 3.68E-40 6.17E-44	J	Palestinian	In data
S10-BC10-P4	America Europe South Asia East Asia Africa	3.36E-36 2.97E-36 1.39E-38 1.38E-41 7.68E-42	J	Palestinian	In data
S11-BC11-P4	America South Asia Europe East Asia Africa	9.19E-37 5.30E-37 9.41E-38 2.92E-39 5.90E-42	J	Jews, Yemenite	In data
S12-BC12-P4	Europe America South Asia East Asia Africa	9.25E-37 7.27E-38 4.47E-38 1.92E-40 1.03E-45	J	Kuwaiti	In data
S13-BC13-P4	America Europe South Asia East Asia Africa	2.85E-39 1.80E-39 3.22E-41 2.40E-41 3.35E-46	J	Jews, Sephardic	In data
S14-BC14-P4	America Europe South Asia East Asia Africa	1.50E-38 4.62E-39 9.27E-40 5.30E-40 8.55E-42	J	Palestinian	Palestinian

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S15-BC15-P4	America Europe South Asia East Asia Africa	5.20E-38 2.66E-39 1.90E-41 1.39E-42 2.68E-04	J	Jews, Ethiopian	Jews, Ethiopian
S16-BC16-P4	Europe America South Asia East Asia Africa	2.71E-38 4.16E-40 8.59E-41 3.02E-43 1.53E-50	J	Palestinian	In data
S17-BC17-P4	South Asia Europe America East Asia Africa	3.32E-35 2.83E-35 3.87E-36 2.55E-38 9.66E-41	J	Keralite	In data
S18-BC18-P4	Europe America South Asia East Asia Africa	1.04E-38 2.84E-39 8.39E-40 3.20E-41 2.73E-46	J	Druze	In data
S19-BC19-P4	Europe America South Asia East Asia Africa	7.64E-37 5.22E-38 4.34E-38 1.86E-38 1.89E-43	J	Kuwaiti	Qatar
S20-BC20-P4	America Europe Africa South Asia East Asia	8.33E-35 4.48E-36 2.97E-37 1.63E-37 2.14E-39	Female	Jews, Ethiopian	In data

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S21-BC21-P4	South Asia Europe America East Asia Africa	2.08E-42 4.44E-43 2.11E-44 2.14E-46 3.89E-50	J	Kuwaiti	In data
S22-BC22-P4	Europe America South Asia East Asia Africa	3.37E-38 1.70E-38 4.40E-40 1.98E-40 2.48E-42	E	Palestinian	In data
S23-BC23-P4	America East Asia Europe South Asia Africa	2.09E-37 1.79E-37 7.50E-38 1.09E-39 2.59E-40	L	Negroid Makrani	In data
S24-BC24-P4	South Asia America Europe East Asia Africa	1.63E-36 4.79E-37 1.62E-37 1.27E-40 1.19E-41	J	Druze	Qatar
S25-BC25-P4	Europe America South Asia East Asia Africa	2.30E-40 2.19E-40 1.01E-41 1.07E-44 6.38E-46	R2	Palestinian	In data
S26-BC26-P4	Europe America South Asia East Asia Africa	1.56E-39 9.46E-40 3.03E-43 2.00E-44 1.81E-50	J	Palestinian	Qatar

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default*	Customized
S27-BC27-P4	East Asia America South Asia Europe Africa	2.08E-39 5.06E-40 4.56E-40 1.03E-40 6.07E-47	Female	Jews, Sephardic	Jews, Sephardic
S28-BC28-P4	Europe South Asia East Asia America Africa	6.55E-39 4.77E-39 1.15E-39 3.90E-40 4.72E-43	Female	Pashtun	Qatar
S29-BC29-P4	Europe America South Asia East Asia Africa	3.34E-40 5.77E-41 1.37E-43 2.72E-48 5.53E-53	R2	Druze	In data
S30-BC30-P4	America Europe South Asia East Asia Africa	3.84E-37 3.42E-37 3.61E-41 1.75E-41 3.97E-46	J	Palestinian	In data

* Default: HID SNP Genotyper default plug-in data set.

** HID SNP Genotyper with Qatar data (samples re-analyzed).

- In data; the sample is included in the data set of the Customized plug-in.

9.16 Appendix 16: Data showing the combined results of the 60 casework samples from Precision ID Ancestry and Identity panel experiments.

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S1-BC1-P5	America	9.43E-34	E	Kuwaiti	Kuwaiti
	Europe	8.96E-34			
	South Asia	1.48E-35			
	Africa	2.02E-36			
	East Asia	4.66E-37			
S2-BC2-P5	America	9.43E-34	E	Kuwaiti	Kuwaiti
	Europe	8.96E-34			
	South Asia	1.48E-35			
	Africa	2.02E-36			
	East Asia	4.66E-37			
S3-BC3-P5	South Asia	1.43E-37	IJ	Not tested	Not tested
	East Asia	7.26E-38			
	Europe	1.15E-39			
	America	2.57E-40			
	Africa	2.39E-42			
S4-BC4-P5	South Asia	1.43E-37	IJ	Not tested	Not tested
	East Asia	7.26E-38			
	Europe	1.15E-39			
	America	2.57E-40			
	Africa	2.39E-42			
S5-BC5-P5	South Asia	3.13E-37	Unknown Female Sample	Hazara	Hazara
	East Asia	3.48E-38			
	America	1.09E-39			
	Africa	1.02E-39			
	Europe	8.32E-41			
S6-BC6-P5	Europe	4.68E-38	J	Palestinian	Palestinian
	America	1.23E-39			
	South Asia	5.48E-42			
	Africa	2.22E-44			
	East Asia	1.60E-45			

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S7-BC7-P5	Europe South Asia America Africa East Asia	3.59E-38 2.54E-39 1.77E-39 4.65E-42 1.19E-43	E	Jews, Yemenite	Qatar
S8-BC8-P5	South Asia America Europe East Asia Africa	8.93E-38 1.83E-38 9.02E-39 2.52E-40 7.46E-45	J	Druze	Druze
S9-BC9-P5	Africa America Europe South Asia East Asia	5.14E-36 5.34E-38 1.13E-38 2.57E-39 1.49E-42	E	Jews, Ethiopian	Jews, Ethiopian
S10-BC10-P5	America Africa South Asia Europe East Asia	4.03E-35 1.43E-35 7.93E-36 3.89E-36 2.08E-37	J	Somali	Somali
S11-BC11-P5	Europe America South Asia East Asia Africa	1.71E-36 4.73E-37 1.94E-37 3.06E-40 6.71E-41	T	Jews, Ethiopian	Jews, Ethiopian
S12-BC12-P5	South Asia East Asia America Europe Africa	1.12E-36 1.45E-37 3.35E-38 6.86E-40 2.96E-42	L	Keralite	Keralite

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S13-BC13-P5	South Asia East Asia America Europe Africa	1.12E-36 1.45E-03 3.35E-38 6.86E-40 2.96E-42	L	Keralite	Keralite
S14-BC14-P5	South Asia Europe America East Asia Africa	2.77E-37 3.09E-38 1.61E-38 2.63E-40 2.85E-41	J	Palestinian	Palestinian
S15-BC15-P5	Europe America South Asia East Asia Africa	5.01E-36 1.88E-36 7.13E-37 2.81E-38 2.49E-39	J	Jews, Yemenite	Jews, Yemenite
S16-BC16-P5	America Europe South Asia East Asia Africa	4.26E-39 3.73E-40 2.38E-42 1.93E-43 1.74E-46	J	Jews, Yemenite	Jews, Yemenite
S17-BC17-P5	South Asia East Asia America Europe Africa	1.12E-36 1.45E-37 3.35E-38 6.86E-40 2.96E-42	L	Keralite	Keralite
S18-BC18-P5	East Asia Europe America South Asia Africa	1.16E-36 3.45E-37 2.75E-37 2.04E-37 3.12E-46	L	Keralite	Keralite

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S19-BC19-P5	Africa South Asia America East Asia Europe	3.92E-39 7.06E-41 8.69E-42 1.29E-42 1.06E-42	E	Negroid Makrani	Negroid Makrani
S20-BC20-P5	East Asia South Asia America Europe Africa	1.57E-37 8.24E-38 1.15E-38 5.38E-39 1.27E-44	J	Kachari	Kachari
S21-BC21-P5	South Asia America Europe East Asia Africa	9.64E-35 5.38E-36 8.71E-37 4.63E-37 3.80E-37	E	Kuwaiti	Kuwaiti
S22-BC22-P5	East Asia South Asia America Europe Africa	1.34E-35 1.49E-38 1.71E-40 1.50E-40 1.38E-46	L	Kachari	Kachari
S23-BC23-P5	South Asia America East Asia Europe Africa	7.01E-38 5.96E-39 3.61E-39 1.58E-40 2.16E-44	J	Jews, Yemenite	Qatar
S24-BC24-P5	Europe America South Asia East Asia Africa	5.33E-38 1.92E-39 3.94E-40 5.60E-42 8.46E-43	J	Jews, Yemenite	Jews, Yemenite

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S25-BC25-P5	Europe America South Asia Africa East Asia	2.51E-39 1.00E-40 8.20E-44 1.56E-44 8.66E-45	J	Jews, Yemenite	Jews, Yemenite
S26-BC26-P5	Europe America South Asia East Asia Africa	3.00E-36 3.51E-37 1.38E-38 1.08E-42 1.96E-47	J	Jews, Yemenite	Qatar
S27-BC27-P5	Europe America South Asia East Asia Africa	4.96E-39 2.98E-39 1.22E-42 2.28E-43 1.46E-46	J	Druze	Qatar
S28-BC28-P5	Europe South Asia America Africa East Asia	1.09E-22 7.53E-23 1.03E-23 2.65E-24 2.54E-26	No haplogroups found.	Not tested	Not tested
S29-BC29-P5	South Asia East Asia America Europe Africa	5.19E-36 3.56E-36 1.04E-36 9.26E-37 2.47E-45	IJ	Hazara	Hazara
S30-BC30-P5	South Asia America Europe East Asia Africa	1.83E-37 3.80E-38 1.28E-38 7.78E-41 2.43E-45	E	Negroid Makran	Negroid Makran

Sample names	Population Names	RMP	Haplogroups	Ancestry Results/ Population likelihood	
				Default	Customized
S1-BC1-P6	South Asia America Europe East Asia Africa	1.77E-37 8.93E-39 7.16E-39 6.01E-40 4.97E-43	R1a1	Keralite	Keralite
S2-BC2-P6	America South Asia Europe East Asia Africa	4.78E-35 5.36E-36 2.14E-36 2.14E-37 4.40E-38	E	Palestinian	Palestinian
S3-BC3-P6	East Asia America South Asia Europe Africa	2.15E-21 7.07E-22 2.28E-22 5.52E-23 1.63E-23	No haplogroups found	Palestinian	Palestinian
S4-BC4-P6	Europe America South Asia Africa East Asia	4.73E-28 5.46E-29 1.69E-29 2.55E-33 2.45E-33	J	Jews, Yemenite	Qatar
S5-BC5-P6					
S6-BC6-P6	East Asia South Asia Europe America Africa	2.06E-02 1.00E-02 3.09E-03 3.06E-03 1.87E-03	No haplogroups found	Hazara	Hazara

PUBLICATIONS

1. Al-Dosari, W. R., Al-Binali, I. A., Pydi, S. S., & Goodwin, W. (2019). Analysis of forensic casework samples by precision ID ancestry panel – manual and automated ampliseq workflow. *Forensic Science International: Genetics Supplement Series*, 7(1), 816-817.

doi:<https://doi.org/10.1016/j.fsigss.2019.10.187>

CONFERENCES/ PROCEEDINGS

1. Poster presentation: Preliminary evaluation of forensic casework samples using the Precision ID Ancestry Panel – manual and automated Ampliseq workflow in the annual Human Identification Solutions HIDS conference, Kobe-Japan, 18-19 June 2019 and in THE 28th OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS, Prague, Czech Republic, 9-13th September 2019.