Human Sample ID Kit

Whole-exome sequencing (WES), whole-genome sequencing (WGS), and gene panel sequencing have become routine in clinical genetic laboratories. The complex workflows, custody transfers, and large datasets impose challenges on data integrity that range from the initial sample collection to the downstream data analysis. As a result, it has been estimated that sample mix-ups occur in up to 3% of the cases, underscoring the need for an independent method for sample identity confirmation.

The Human Sample ID Kit is an easy and flexible resequencing assay that targets 44 highly polymorphic SNPs and 6 gender markers, creating a unique genetic label for each sample. Using a simple closed-tube PCR step, sequencing-ready libraries are generated that are compatible with Illumina sequencing instruments. The Human Sample ID Kit is designed to be compatible with established target enrichment platforms and is validated for high-quality genomic, low-quality formalin-fixed paraffin-embedded (FFPE), and cell-free DNA (cfDNA) samples.

Advantages

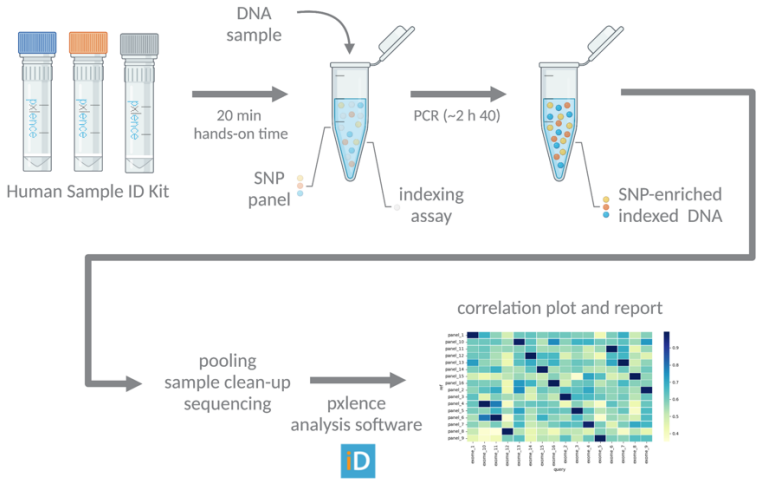
- Superior coverage uniformity
- Easy integration in WES and WGS workflows
- Simple and flexible workflow
- Low template input, down to 2 ng DNA
- Analysis software and standardized report

Key Features

- Works with high-quality, FFPE and liquid biopsy cell-free DNA samples
- Very high discrimination power >1:1000,000,000,000,000,000
- Ambient temperature kit shipment
- Maximum compatibility with established enrichment platforms

Workflow

The protocol consists of a single PCR step simultaneously amplifying the 50 targets and indexing with sample-specific barcodes (dual indexes). After a simple clean-up, the libraries are ready to be sequenced. An automated data analysis report is provided by our accompanying analysis software.



Performance

Uniform coverage across different types and input amounts of DNA

The Human Sample ID Kit was applied on 6 different Genome in a Bottle (GiB) reference samples of the Ashkenazim and Asian trios (Coriell), a synthetic cfDNA sample (SensID), and DNA isolated from FFPE tissues.

Uniform coverage is obtained for all samples, with 88.6% of the SNPs having a **coverage within 2-fold of the median** for the high-quality GiB samples, and 80.3 % and 79.5 % for the cfDNA and FFPE samples, respectively (Figure 1). Data shows that the degree of coverage uniformity is independent from the DNA input amount. All 44 polymorphic SNPs show a coverage above 20% of the mean coverage. Only at very low DNA input amounts (1 - 5 ng) of the cfDNA and FFPE samples, a single SNP occasionally dropped slightly below the 20 % of the mean coverage threshold (Figure 1).

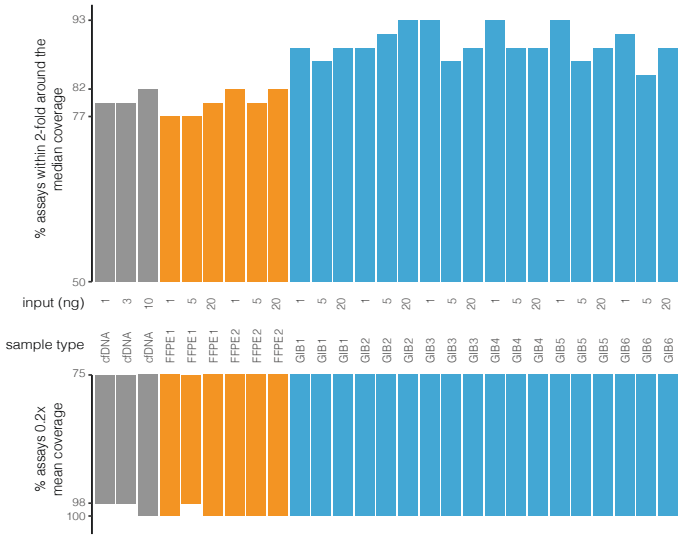


Figure 1

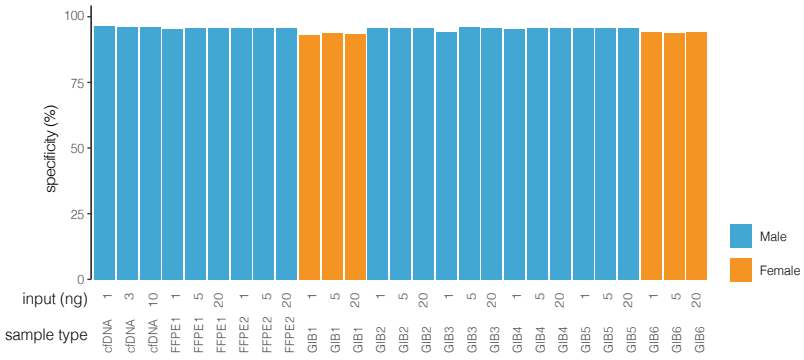


Figure 2

Highly target specific

The Human Sample ID Kit shows a mean on-target mapping rate of 95.3% across all loci. The small difference between male (95.7%) and female (93.8%) DNA samples is related to the gender markers included in the Kit (Figure 2).

High-confidant genotyping and identification

The Human Sample ID Kit allows high-confident variant calling for all SNPs in all samples, resulting in unambiguous sample discrimination and identification using available genotype information for the GiB and cfDNA samples (Figure 3). The correlation heatmap clearly shows that the Kit can elucidate the genetic relationship among the family members of two trios, while providing sufficient discrimination power to distinguish the individual family members. Hassle-free results with the accompanying data analysis software.

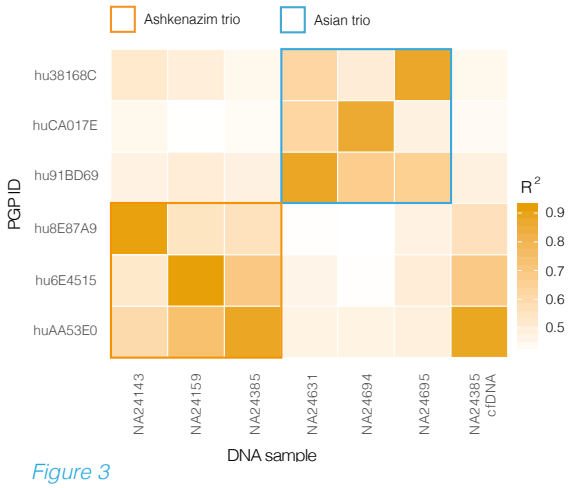


Figure 3

Highlights

Feature	Specification
Sample type	fresh/frozen tissues/cells, fixed tissues, liquid biopsies
Input requirements	1 – 20 ng
Total time	3 h
Hands-on time	20 min
Number of loci	50 (incl. 6 gender markers)
Discrimination power	> 1:1000,000,000,000,000,000 (<1E-16)
Amplicon length	61 – 101 bp
On-target percentage	> 90 %
Coverage uniformity, > 20% of mean	100 % of the loci
Compatible platform	Illumina
Number of samples per million read pairs*	> 96 samples

* Pooled libraries can be sequenced separately or spiked in exome or whole genome sequencing libraries.

Maximum compatibility with exome enrichment methods

The 44 SNPs in the Human Sample ID Kit are carefully selected to be fully compatible with the following commonly used exome enrichment platforms.

Manufacturer	Enrichment kit
Agilent	SureSelect Human All Exon V5 [+UTR]
	SureSelect Human All Exon V6 [+UTR]
	SureSelect Human All Exon V7
	SureSelect Human All Exon V8
	SureSelect Clinical Research Exome [+ v2]
	SureSelectXT Human All Exon V6+COSMIC
	SureSelect Focused Exome
	SureSelectXT Human All Exon V8
	ClearSeq Inherited Disease Panel
Roche	SeqCap EZ Exome + UTR
	SeqCap EZ MedExome
	SeqCap EZ HGSC VCRome
	SeqCap EZ Prime Exome
	SeqCap EZ Inherited Disease Panel
	KAPA HyperExome
	KAPA HyperExome V2
Illumina	TruSeq DNA Exome
	AmpliSeq Exome Panel
	TruSight One Sequencing Panel [+ expanded]
IDT	xGen Exome Research Panel v1
	xGen Exome Research Panel v2
Twist Bioscience	Human Core Exome Panel
	Human RefSeq Panel

Human Sample ID targets

The 44 SNPs[1] in the Human Sample ID Kit are carefully selected to have high minor allele frequencies across a wide range of populations to ensure maximal discrimination potential. In addition, 6 gender markers are included. Most SNPs originate from Pengelly *et al.*, Genome Medicine, 2013.

Chr	Position[2]	Reference ID	REF	ALT
chr1	179551371	rs1410592	G	A
chr1	209638541	rs2076356	T	G
chr1	209795339	rs2013162	C	A
chr2	169235885	rs2229267	A	G
chr2	178589667	rs1560221	A	G
chr2	178590480	rs2163009	T	C
chr2	214955289	rs10498027	G	A
chr2	44275649	rs3738985	A	C
chr4	5748177	rs4688963	T	C
chr4	87613083	rs2736982	A	G
chr5	136056737	rs4669	T	C
chr5	139121126	rs3088052	T	C
chr5	54456158	rs7823	T	C
chr7	33970334	rs10265207	C	T
chr7	43807004	rs7738	A	G
chr8	103324868	rs3808554	A	G
chr9	133439376	rs3124768	A	G
chr9	27202872	rs639225	A	G
chr9	74800368	rs7859201	A	C
chr10	102837167	rs6163	C	A
chr10	68166340	rs2673794	T	C
chr10	77184832	rs1131824	G	A
chr10	94032006	rs17109674	G	A
chr11	6608435	rs1043388	C	T
chr12	1806958	rs60637	C	A
chr12	884764	rs7300444	C	T

Chr	Position[2]	Reference ID	REF	ALT
chr13	24892817	rs3742165		C
chr13	38859469	rs9532292	A	G
chr14	64170429	rs7161192	C	A
chr16	68679827	rs2296409	G	A
chr16	68679920	rs2296408	C	A
chr16	68695882	rs17715450	C	A
chr16	70512331	rs3762171	G	A
chr17	10632701	rs2285479	G	A
chr17	10639154	rs2285475	T	G
chr17	44372421	rs5910	G	A
chr17	73196524	rs1052706	G	A
chr18	23833905	rs9962023	T	A
chr18	49929553	rs2298628	C	T
chr19	10156401	rs2228611	T	C
chr19	32862558	rs11084673	G	A
chr20	6119441	rs10373	A	G
chr21	46353189	rs2249057	C	A
chr22	37073551	rs4820268	G	A
chrX/Y		AMELX/Y		
chrY		KDM5D		
chrY		SRY		
chrY		TXLNGY		
chrY		USP9Y		
chrY		UTY		

[1] Method for selecting SNPs based upon Pengelly *et al.* A SNP profiling panel for sample tracking in whole-exome sequencing studies. *Genome Med* **5,** 89 (2013). https://doi.org/10.1186/gm492

[2] Positions according to the human reference genome GRCh38, hg38

Discrimination potential

The Human Sample ID Kit has extremely high discrimination potential across all tested human populations. The discrimination potential is the chance that 2 random individuals share the same SNP genotype, also called the random match probability.

Population	Discrimination potential[3]
African / African-American	7.29E-18
Amish	1.36E-17
Ashkenazi Jewish	2.32E-18
East Asian	4.63E-18
European (Finnish)	3.88E-18
European (non-Finnish)	3.97E-18

Population	Discrimination potential[3]
Latino / admixed American	1.89E-18
Middle Eastern	3.93E-18
South Asian	2.41E-18
other	1.41E-18
overall	9.40E-19

[3] Discrimination potential calculated as a function of the allele frequencies from gnomAD v3.1.1.

Minor allele frequencies for different populations[4]

Reference ID	African / African-American	Amish	Ashkenazi Jewish	East Asian	European (Finnish)	European (non-Finnish)	Latino / admixed American	Middle Eastern	South Asian	other	overall
rs1410592	0.40	0.35	0.36	0.47	0.38	0.37	0.36	0.40	0.39	0.41	0.39
rs2076356	0.42	0.25	0.33	0.28	0.28	0.35	0.50	0.28	0.43	0.48	0.44
rs2013162	0.43	0.49	0.43	0.39	0.37	0.37	0.36	0.35	0.32	0.28	0.37
rs2229267	0.35	0.23	0.35	0.41	0.19	0.24	0.44	0.39	0.39	0.37	0.38
rs1560221	0.30	0.41	0.30	0.47	0.50	0.35	0.37	0.38	0.40	0.41	0.46
rs2163009	0.29	0.44	0.46	0.48	0.39	0.33	0.29	0.29	0.28	0.23	0.36
rs10498027	0.31	0.46	0.49	0.24	0.41	0.42	0.41	0.47	0.42	0.46	0.39
rs3738985	0.20	0.29	0.20	0.36	0.33	0.23	0.31	0.19	0.25	0.16	0.25
rs4688963	0.47	0.44	0.35	0.37	0.41	0.37	0.35	0.30	0.39	0.47	0.41
rs2736982	0.27	0.47	0.44	0.31	0.32	0.39	0.42	0.27	0.36	0.41	0.35
rs4669	0.38	0.24	0.30	0.33	0.29	0.27	0.33	0.24	0.37	0.37	0.38
rs3088052	0.33	0.35	0.49	0.17	0.45	0.44	0.41	0.50	0.47	0.37	0.49
rs7823	0.27	0.28	0.41	0.36	0.37	0.38	0.49	0.32	0.37	0.43	0.37
rs10265207	0.39	0.48	0.43	0.43	0.36	0.47	0.39	0.42	0.42	0.47	0.43
rs7738	0.45	0.39	0.47	0.41	0.38	0.37	0.39	0.48	0.43	0.45	0.43
rs3808554	0.31	0.30	0.36	0.39	0.44	0.47	0.48	0.38	0.43	0.44	0.43
rs3124768	0.36	0.44	0.37	0.19	0.37	0.43	0.50	0.43	0.48	0.48	0.44
rs639225	0.37	0.50	0.37	0.40	0.36	0.49	0.49	0.40	0.43	0.38	0.46
rs7859201	0.33	0.35	0.37	0.36	0.41	0.38	0.35	0.34	0.40	0.41	0.46
rs6163	0.37	0.20	0.40	0.44	0.35	0.40	0.40	0.37	0.39	0.35	0.39
rs2673794	0.35	0.34	0.36	0.30	0.25	0.31	0.36	0.50	0.41	0.39	0.42
rs1131824	0.37	0.30	0.41	0.26	0.39	0.35	0.32	0.35	0.38	0.33	0.43
rs17109674	0.48	0.22	0.25	0.44	0.31	0.25	0.23	0.25	0.29	0.32	0.32
rs1043388	0.38	0.40	0.30	0.35	0.24	0.26	0.26	0.32	0.29	0.24	0.29
rs60637	0.31	0.18	0.34	0.43	0.29	0.24	0.29	0.40	0.35	0.37	0.40
rs7300444	0.31	0.49	0.39	0.47	0.44	0.44	0.42	0.27	0.38	0.35	0.40
rs3742165	0.30	0.31	0.49	0.28	0.37	0.47	0.42	0.46	0.49	0.49	0.48
rs9532292	0.45	0.31	0.41	0.38	0.25	0.31	0.39	0.41	0.36	0.45	0.39
rs7161192	0.35	0.29	0.38	0.43	0.34	0.32	0.29	0.31	0.32	0.29	0.34
rs2296409	0.33	0.28	0.43	0.37	0.39	0.37	0.46	0.35	0.38	0.45	0.37
rs2296408	0.33	0.28	0.43	0.37	0.39	0.37	0.46	0.35	0.38	0.46	0.37
rs17715450	0.31	0.31	0.45	0.50	0.40	0.41	0.48	0.42	0.49	0.46	0.50
rs3762171	0.29	0.36	0.43	0.43	0.35	0.42	0.45	0.50	0.42	0.35	0.39
rs2285479	0.30	0.25	0.30	0.37	0.29	0.27	0.44	0.45	0.41	0.48	0.43
rs2285475	0.30	0.26	0.30	0.37	0.29	0.27	0.44	0.45	0.41	0.48	0.43
rs5910	0.39	0.41	0.44	0.44	0.44	0.38	0.35	0.32	0.36	0.35	0.39

Reference ID	African / African-American	Amish	Ashkenazi Jewish	East Asian	European (Finnish)	European (non-Finnish)	Latino / Admixed American	Middle Eastern	South Asian	Other	Overall
rs1052706	0.27	0.49	0.44	0.40	0.48	0.47	0.46	0.49	0.47	0.47	0.44
rs9962023	0.46	0.23	0.22	0.25	0.36	0.30	0.40	0.21	0.29	0.37	0.36
rs2298628	0.41	0.27	0.41	0.25	0.48	0.44	0.41	0.44	0.50	0.46	0.49
rs2228611	0.49	0.44	0.44	0.33	0.48	0.50	0.49	0.45	0.49	0.45	0.49
rs11084673	0.31	0.38	0.33	0.29	0.39	0.34	0.23	0.38	0.32	0.21	0.32
rs10373	0.46	0.43	0.39	0.38	0.41	0.47	0.50	0.36	0.45	0.36	0.48
rs2249057	0.23	0.38	0.33	0.40	0.31	0.33	0.26	0.41	0.31	0.31	0.30
rs4820268	0.27	0.40	0.45	0.47	0.44	0.46	0.48	0.46	0.45	0.43	0.42

[4] Minor allele frequencies (MAFs) obtained from gnomAD v3.1.1.

Ordering Information

Product name	Catalogue number
Human Sample ID Kit (96 rxn)	PXL-SID-001
Human Sample ID CDI Set (12 x 8; 96 rxn)*	PXL-IND-001
Human Sample ID UDI Set 1 (2 x 96 rxn)*	PXL-IND-002

* For more indexing options contact info@pxlence.com.

Visit www.pxlence.com for ordering.

For research use only.

pxlence bv
Stationsstraat 30a
B-9200 Dendermonde
Belgium
info@pxlence.com