HW1 – NGS Workflows for Forensic Genetics
Type: Half Day Workshop, Lecture
Date & Time: Monday, 9 September (9:00-13:00)

Peter M. Vallone, Ph.D.
Leader, Applied Genetics Group NIST
September 9, 2019
Prague, Czech Republic

Our day together (4 hours)

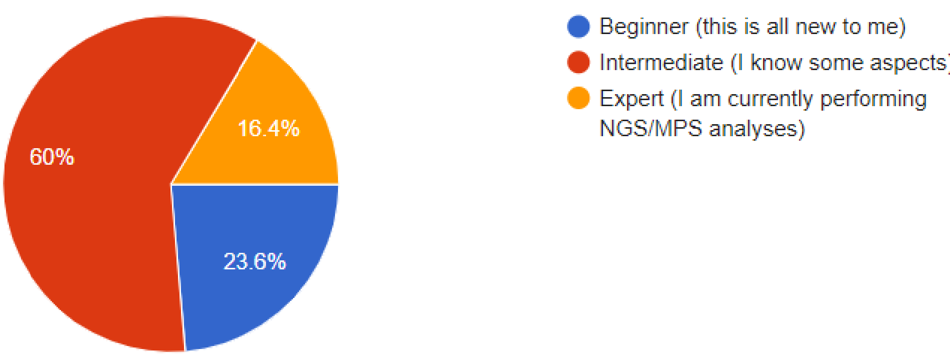
- 9:00 – 10:30 AM: Introduction & Workflow 1
- 10:30 – 11:00 AM: Coffee Break!
- 11:00 AM – 1:00 PM: Workflows 2 & 3

Slides and supporting information can be found at:
Google drive link
https://drive.google.com/drive/folders/1cbqsnBaEgivRkYIxNr2r2LtBFibvafvk
Or use
https://tinyurl.com/yyfwe8jc

This information was sent out by email prior to the workshop

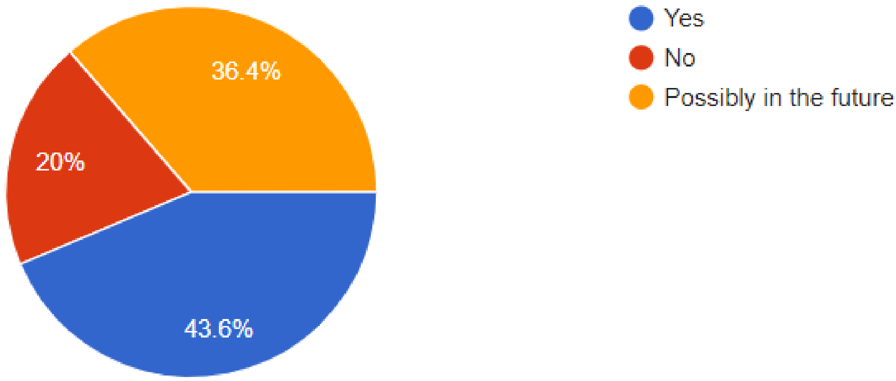
What is your personal level of experience with NGS/MPS methods?

55 responses



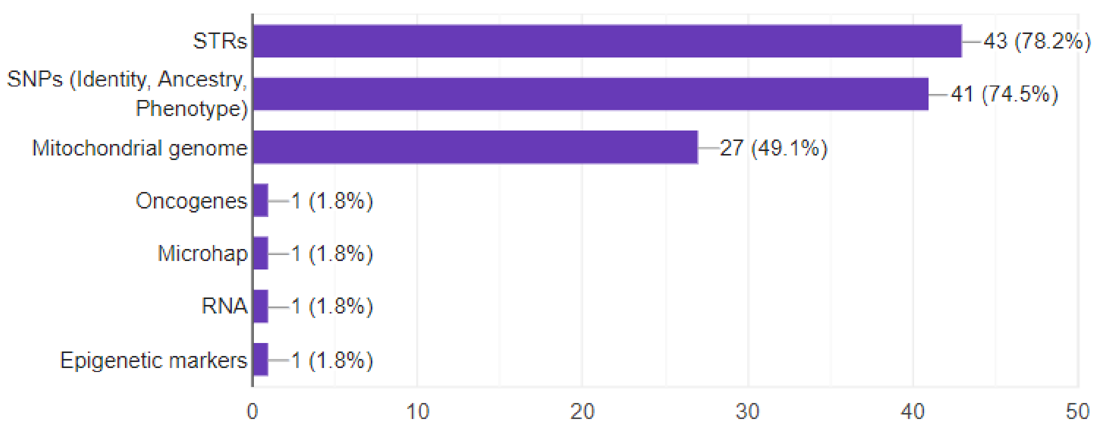
Is your lab currently performing NGS/MPS analyses?

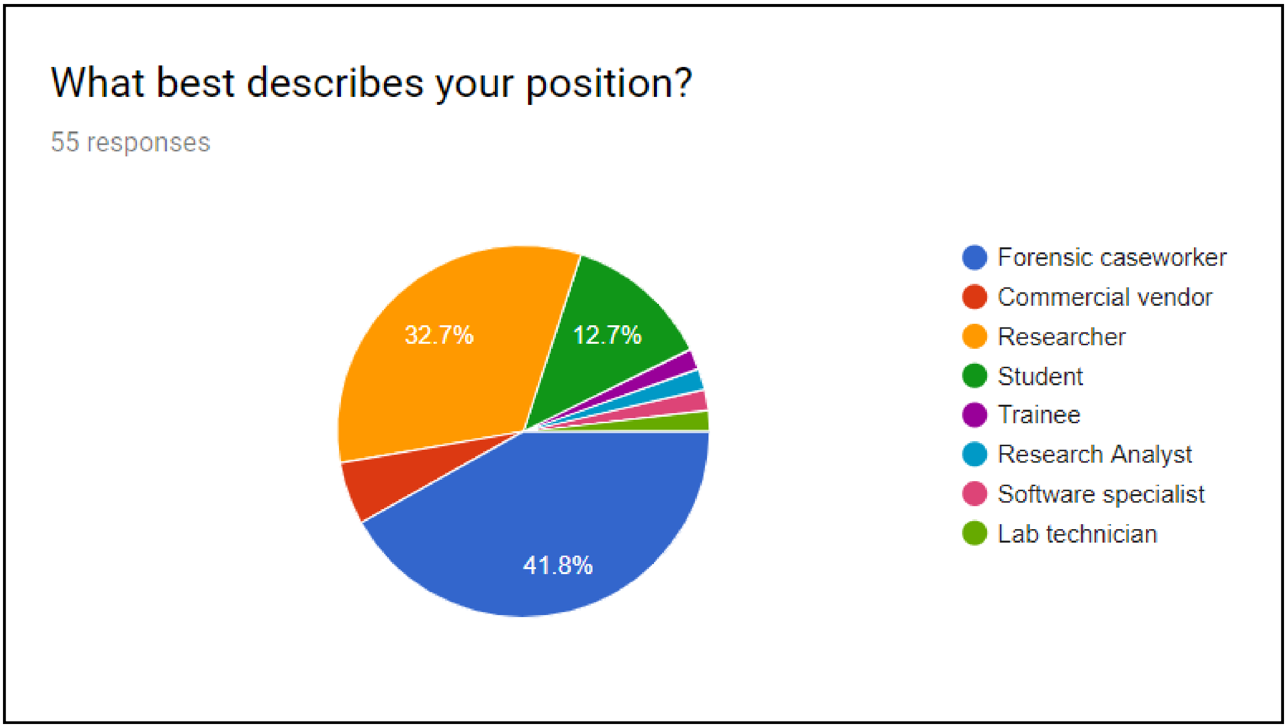
55 responses



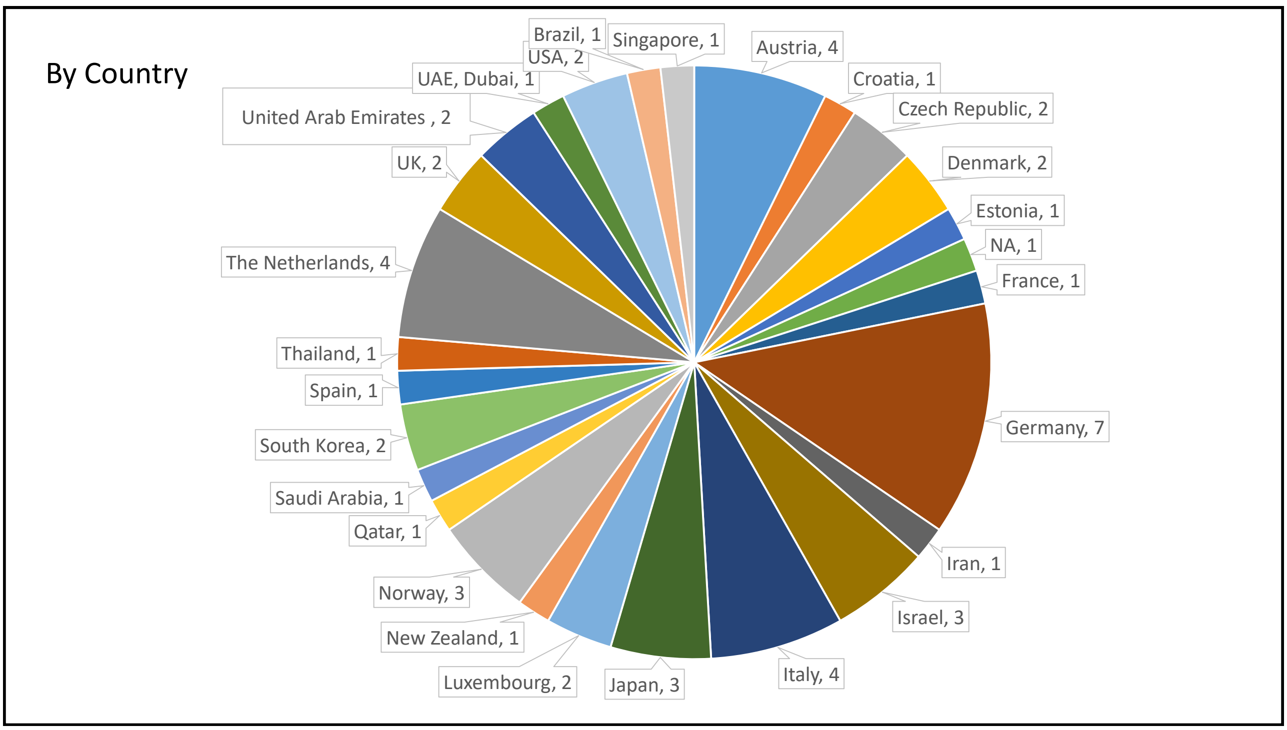
What marker systems are you interested in sequencing? (you can select more than one)

55 responses





What best describes your position?
55 responses
32.7%
12.7%
41.8%
Forensic caseworker
Commercial vendor
Researcher
Student
Trainee
Research Analyst
Software specialist
Lab technician



By Country
Brazil, 1
Singapore, 1
USA, 2
Austria, 4
UAE, Dubai, 1
Croatia, 1
United Arab Emirates , 2
Czech Republic, 2
UK, 2
Denmark, 2
Estonia, 1
NA, 1
The Netherlands, 4
France, 1
Thailand, 1
Spain, 1
Germany, 7
South Korea, 2
Saudi Arabia, 1
Qatar, 1
Iran, 1
Norway, 3
Israel, 3
New Zealand, 1
Italy, 4
Luxembourg, 2
Japan, 3

- There is an interest in sequencing as a forensic workflow
  - PCR-Sequencing
- Differs from traditional PCR-CE workflow
  - How?
  - What is the same?
  - What is different?

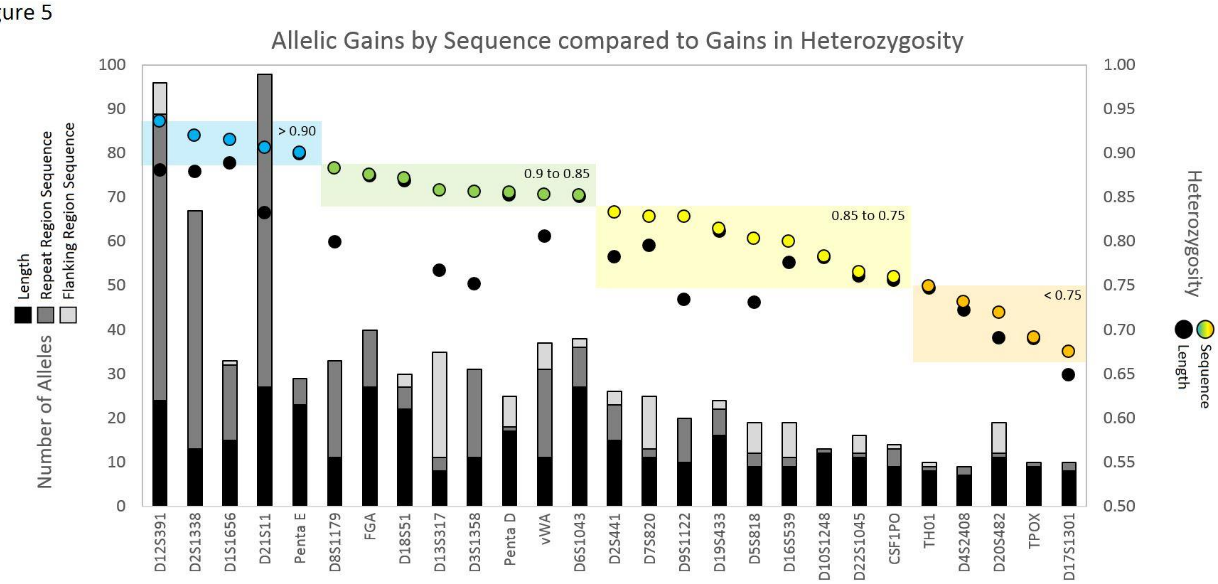
Why sequence?

- More markers – higher multiplexing capability (versus CE-based methods)
- More information – more markers and sequence level resolution
- **End goal:** Access to this additional information will support forensic casework applications

Why sequence STRs?

- Sequencing of STRs
  - STR motif sequence variation; flanking region variation (more polymorphic)
  - Further understand simple versus complex repeat motifs
  - Characterize stutter
- Applications
  - One to one matching? – RMPs with 20 STR markers are already quite low ($>10^{-20}$)
  - Partial profiles
  - Kinship
  - Length-based allele calls are back compatible with existing databases
- Mixtures
  - Resolve alleles identical by length, but differ by sequence
  - Separate stutter from low level contributors (based on sequence)
  - A sequenced allele *may* have a lower frequency (resulting in a higher LR)

Figure 5



Gettings et al. (2018) Sequence-based US population data for 27 autosomal STR loci. Forensic Sci Int Genet 37:106-115.

Sequencing other marker systems: applications

- Mitochondrial DNA sequencing
  - Control region and/or full genome

Higher throughput than Sanger methods
Measure lower levels of heteroplasmy
Easier workflow ?

- SNPs
  - Ancestry, Identity, Phenotype, Microhaplotypes (closely linked SNPs)
- In the future?
  - Non-targeted sequencing, RNA targets, metagenomics, epigenetics
  - New technologies, methods, marker systems

This week
Phenotype-Ancestry
Single cell
Methylation – Age prediction
Mixtures
Casework
Microbiome
RNA

- O003: DEVELOPMENT AND OPTIMIZATION OF THE VISAGE PROTOTYPE TOOLS FOR BIO-GEOGRAPHIC ANCESTRY AND APPEARANCE TRAITS INFERENCE USING TARGETED MPS
- O007: COMPARISON OF CE- AND MPS-BASED ANALYSES OF FORENSIC MARKERS WITH SINGLE CELL AFTER WHOLE GENOME AMPLIFICATION
- O008: PRESENTATION OF THE HUMAN PIGMENTATION (HUPI) AMPLISEQ CUSTOM PANEL
- O009: PREDICTIVE DNA ANALYSIS OF HUMAN HEAD HAIR GREYING USING WHOLE-EXOME AND TARGETED NGS DATA EXAMINED WITH DEEP LEARNING METHODS
- O010: A COMPARISON OF DNA METHYLATION TECHNOLOGIES AND PERFORMANCE OF AGE PREDICTION MODELS
- O015: COMPARISON OF CE AND MPS BASED ANALYSIS FOR THE PROBABILISTIC INTERPRETATION OF MIXED STR PROFILES
- O017: THE FIRST MPS-STR BASED CONVICTION IN A CRIMINAL CASE?
- O018: ENHANCING STR SEQUENCE ALLELE REPRESENTATION FOR PROBABILISTIC GENOTYPING
- O019: A MASSIVELY PARALLEL SEQUENCING ASSAY OF MICROHAPLOTYPES FOR MIXTURE DECONVOLUTION
- O059: TAXONOMY-INDEPENDENT DEEP LEARNING MICROBIOME APPROACH FOR ACCURATE CLASSIFICATION OF FORENSICALLY RELEVANT HUMAN BIOMATERIALS USING TARGETED MPS
- O060: PERFORMANCE OF ENVIRONMENTAL DNA METABARCODING IN SOIL TRACE MATCHING AND PROVENANCING
- O063: SPECIES IDENTIFICATION USING MASSIVELY PARALLEL SEQUENCING – DETECTING MULTIPLE SPECIES IN MIXED SOURCES
- O065: WHOLE-GENOME SEQUENCING OF NEISSERIA GONORRHOEAE IN A FORENSIC TRANSMISSION CASE
- O066: WHOLE TRANSCRIPTOME ANALYSIS OF AGED BIOLOGICAL CRIME SCENE TRACES

Additional related workshops…

HW8 - Autosomal STR Genomics 101: Sequence Variation and Nomenclature

Organiser: Katherine Gettings
Date & Time: Tuesday, 10 September (14:00-18:00)
Capacity: 60pax
Workshop type: Lecture

HW9 - Forensic DNA Phenotyping: basics of data acquisition and interpretation

Organiser: Wojciech Branicki
Date & Time: Tuesday, 10 September (14:00-18:00)
Capacity: 100pax
Workshop type: Lecture

FW3 - Population analysis of forensic DNA data using Snipper and STRUCTURE

Organiser: Christopher Phillips & Leonor Gusmao
Date & Time: Monday, 9 September (14:00-18:00) & Tuesday, 10 September (9:00-13:00)
Capacity: 40pax
Workshop type: Hands-on

Targeted Sequencing

- We wish to sequence 'our' markers (STRs, SNPs, mtGenome)
- You might 'see' these markers in whole genome sequencing (shotgun)
  - Low coverage
  - Issues with STR regions
  - Can be more bioinformatically challenging
  - Inefficient use of the sequencing 'space'

And would require more sample 100 ng – 1 µg

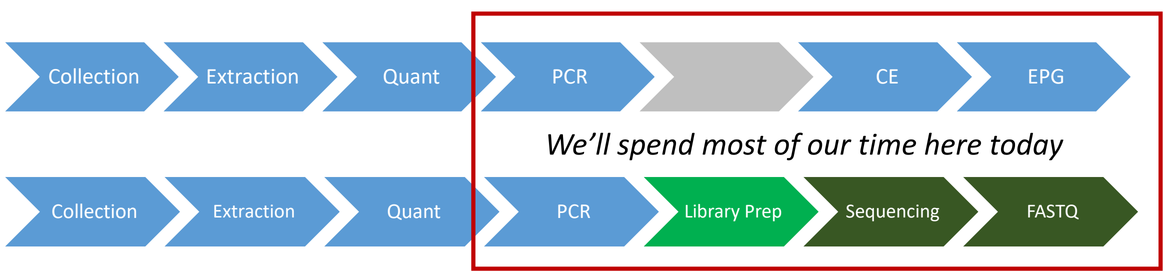
- For now…it seems that forensic genetics will implement a targeted approach

Is Sequencing more sensitive?

- More sensitive than what? CE-based testing?
- Targeted sequencing is PCR-based
  - Still using PCR, stochastic effects are encountered at low amounts of DNA
  - Expect similar levels of sensitivity (in terms of generating PCR products)
- 'Sensitive' can also relate to: more markers, more information, improved matching statistics
- Sequencing methods *may* allow for a deeper understanding of S/N and artifacts

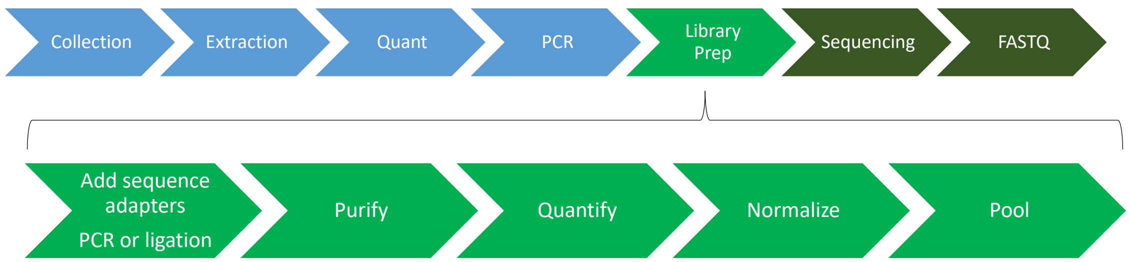
Comparing workflows – targeted sequencing

PCR-Capillary Electrophoresis



PCR-Sequencing

Library preparation



General…we'll get into specifics today

Goal: prepare PCR products for sequencing

Outline for today

We will discuss various sequencing workflows

1. Sequencing method: targeted PCR & library preparation
2. Sequencing platform
3. Data analysis: brief examples

Select listing of commercial sequencing workflows

Assay	Platform	Associated Software	Markers
ForenSeq DNA Signature Prep Kit	MiSeq FGx	UAS	auSTRs, Y STRs, X STRs and SNPs
ForenSeq mtDNA Control Region Solution	MiSeq FGx	UAS	Mitochondrial control region (WG soon?)
PowerSeq 46GY System	MiSeq	Open	auSTR and Y STRs
PowerSeq CRM Nested System, Custom	MiSeq	Open	Mitochondrial control region (and WG)
Precision ID SNP Identity Panel	S5	Converge	Identity SNPs
Precision ID SNP Ancestry Panel	S5	Converge	Ancestry SNPs
Precision ID STR GlobalFiler NGS STR Panel v2	S5	Converge	Autosomal STRs
Precision ID mtDNA Whole Genome Panel	S5	Converge	Whole mitochondrial genome
Precision ID mtDNA Control Region Panel	S5	Converge	Mitochondrial control region
Precision ID SNP Phenotype Panel	S5	Converge	SNPs
GeneReader DNAseq Targeted Panels V2	Illumina/S5	CLCBio - open	Mito, SNPs

UAS = Universal Analysis Software

List not exhaustive – just some common examples

Focus of the NGS workflows for today

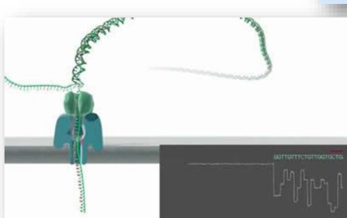
	assay	instrument	software	markers
Verogen/Illumina	ForenSeq DNA Signature Prep Kit	MiSeq FGx	Universal Analysis Software (UAS)	STRs, SNPs
Promega/Illumina	PowerSeq 46 GY System	MiSeq	STRait Razor toaSTR	STRs
Thermo Fisher	Precision ID Panels	S5	Converge	STRs, SNPs

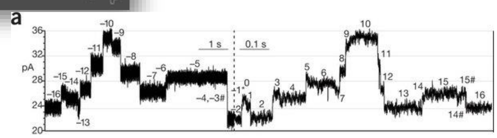
Time permitting we will also discuss mitochondrial control region and whole mito genome sequencing

Additional Sequencing Technologies

- QIAGEN GeneReader
  - Sequencing by synthesis approach
  - QIACube NGS (for automated library preparation)
  - QIAGEN has also purchased CLC bio

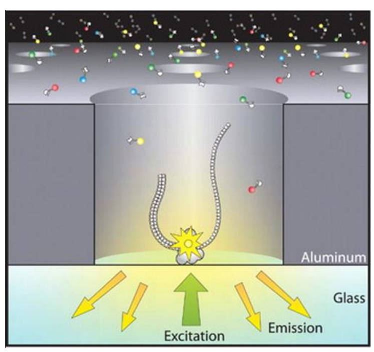
- Oxford Nanopore (MinION)
  - Ratcheting strand of DNA through a protein manifold
  - Bases are detected by a difference in current





Additional Sequencing Technologies

- PacBio
- Bases are incorporated and detected in real-time
- No PCR needed
- Long reads - 1000s of bases



https://www.pacb.com/wp-content/uploads/Infographic_SMRT-Sequencing-How-it-Works.pdf

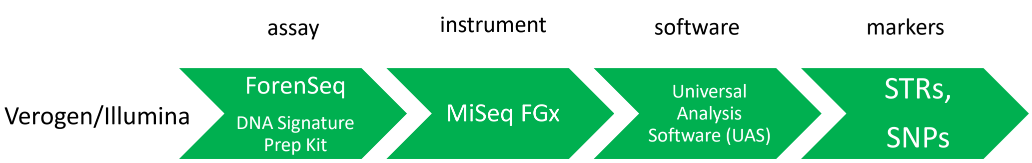
Disclaimer

- *I am not covering all possible workflows that might apply to forensic genetics*
- *I hope that the ones we discuss to today are useful to illustrate the methods and techniques you will encounter*
- *Not my intention to suggest that one method or platform is best*
- *It is up to you to decide want you need (cost, time, steps, automation, information, ease of use, throughput, data, marker systems, etc.)*
- *More information can be found in the manuals and specific literature*

Acknowledgments

- Verogen
  - John Walsh
  - Meghan Didier
  - Melissa Kotkin
- Thermo Fisher
  - Matt Gabriel
  - Chantal Roth
- toaSTR
  - Sebastian Ganschow
- NIST
  - Kevin Kiesler
  - Sarah Riman
  - Katherine Gettings
  - Lisa Borsuk

NGS workflow #1



ForenSeq™ DNA Signature Prep kit - MiSeq FGx

Table 1: ForenSeq DNA Signature Prep Kit—Forensic Loci

Feature	Number of Markers[a]	Amplicon Size Range (bp)	Included in DNA Primer Mix A	Included in DNA Primer Mix B[b]
Global Autosomal STRs	27	61–467	Yes	Yes
Y-STRs	24	119–390	Yes	Yes
X-STRs	7	157–462	Yes	Yes
Identity SNPs	94	63–231	Yes	Yes
Phenotypic SNPs	22	73–227	No	Yes
Biogeographical Ancestry SNPs	56	67–200	No	Yes

a. SNP and STR chromosome locations can be found in the ForenSeq DNA Signature Prep Kit Reference Guide.
b. Over 200 markers analyzed when running primer set B.

Jäger AC, Alvarez ML, Davis CP, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for targeted next generation sequencing in forensic DNA casework and database laboratories. Forensic Sci Int. 2017;28: 52-70

What is the overall goal of library preparation?

- To prepare the PCR products for the sequencer
- Capture a ‘snapshot’ of the PCR products (ratios, abundance)
- We want to avoid
  - Any bias that favors a product based on size, sequence, abundance
  - Uneven yields or representation across samples
  - Inefficient use of the sequencing capability

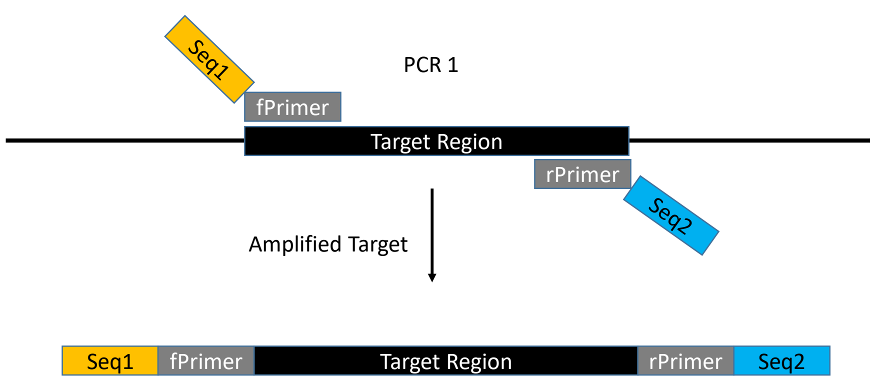
ForenSeq (Sequenced on MiSeq FGx)

Targeted PCR (tags added) → Enrich Targets (indices added) → Purify libraries → Normalize libraries → Pool libraries → Denature and prepare for loading

---

Workflow: ForenSeq™ DNA Signature Prep kit
Step: Targeted PCR (PCR 1)
Purpose: Targeted amplification of STRs/SNPs

Target 1 ng of gDNA
18 PCR cycles
5 µL of DNA extract
15 µL total volume

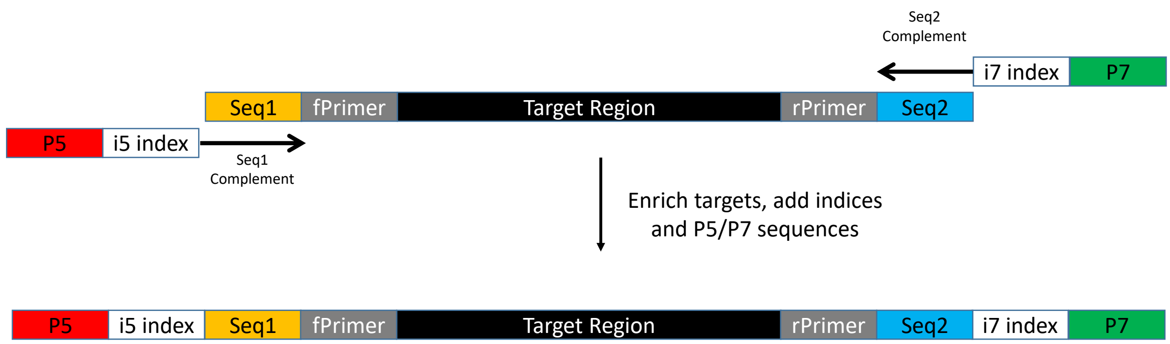


Workflow: ForenSeq™ DNA Signature Prep kit

Step: Enrich targets (PCR 2)

Purpose: Add indices and P5/P7 sequences to the PCR products

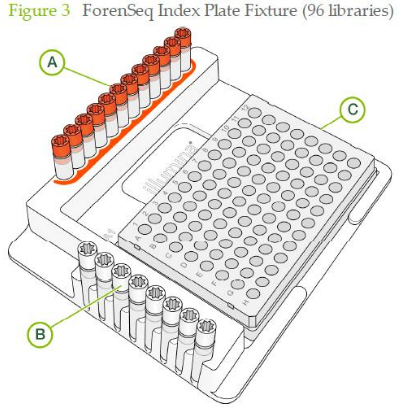
PCR 1 products
15 µL volume
15 PCR cycles
Perform in a different room
(not in the original PCR 1 environment)



Index Adapters

P5: 5' AAT GAT ACG GCG ACC ACC GA 3’
P7: 5' CAA GCA GAA GAC GGC ATA CGA GAT 3'

Figure 3 ForenSeq Index Plate Fixture (96 libraries)



A Columns 1–12: Index 1 (i7) adapters (orange caps)
B Rows A–H: Index 2 (i5) adapters (white caps)
C FSP plate

i5 index name		i7 index name	
A501	TGAACCTT	R701	ATCACG
A502	TGCTAAGT	R702	CGATGT
A503	TGTTCTCT	R703	TTAGGC
A504	TAAGACAC	R704	TGACCA
A505	CTAATCGA	R705	ACAGTG
A506	CTAGAACA	R706	GCCAAT
A507	TAAGTTCC	R707	CAGATC
A508	TAGACCTA	R708	ACTTGA
		R709	GATCAG
		R710	TAGCTT
		R711	GGCTAC
		R712	CTTGTA

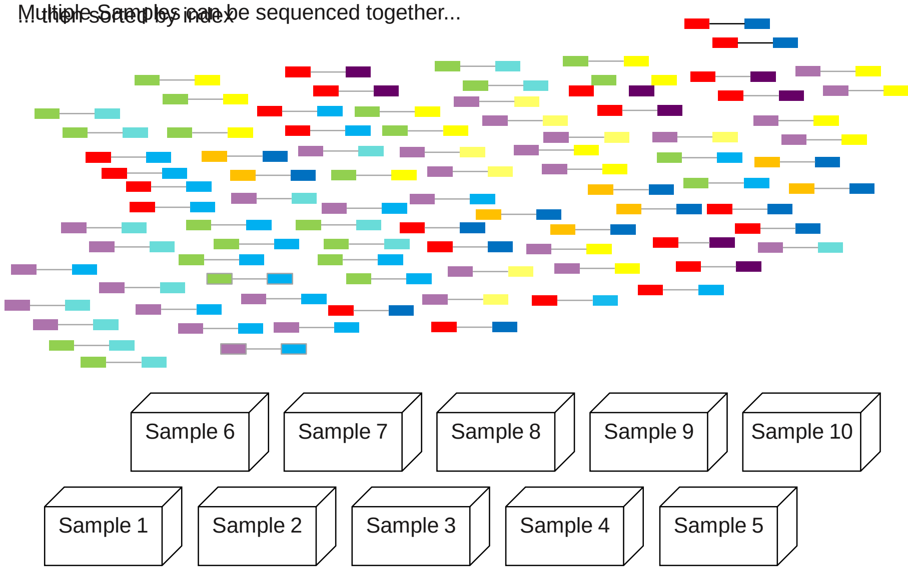
96 unique combination can be created from the eight i5 and twelve i7 indices

	R701	R702	R703	R704	R705	R706	R707	R708	R709	R710	R711	R712
A501	A501 R701	A501 R702	A501 R703	A501 R704	A501 R705	A501 R706	A501 R707	A501 R708	A501 R709	A501 R710	A501 R711	A501 R712
A502	A502 R701	A502 R702	A502 R703	A502 R704	A502 R705	A502 R706	A502 R707	A502 R708	A502 R709	A502 R710	A502 R711	A502 R712
A503	A503 R701	A503 R702	A503 R703	A503 R704	A503 R705	A503 R706	A503 R707	A503 R708	A503 R709	A503 R710	A503 R711	A503 R712
A504	A504 R701	A504 R702	A504 R703	A504 R704	A504 R705	A504 R706	A504 R707	A504 R708	A504 R709	A504 R710	A504 R711	A504 R712
A505	A505 R701	A505 R702	A505 R703	A505 R704	A505 R705	A505 R706	A505 R707	A505 R708	A505 R709	A505 R710	A505 R711	A505 R712
A506	A506 R701	A506 R702	A506 R703	A506 R704	A506 R705	A506 R706	A506 R707	A506 R708	A506 R709	A506 R710	A506 R711	A506 R712
A507	A507 R701	A507 R702	A507 R703	A507 R704	A507 R705	A507 R706	A507 R707	A507 R708	A507 R709	A507 R710	A507 R711	A507 R712
A508	A508 R701	A508 R702	A508 R703	A508 R704	A508 R705	A508 R706	A508 R707	A508 R708	A508 R709	A508 R710	A508 R711	A508 R712

Sample Demultiplexing Using Index Sequences

Multiple Samples can be sequenced together...

...then sorted by index



Slide courtesy of Meghan Didier (Verogen)

Combinations of i5 (8 nt) and i7 (6 nt) will allow for sample barcoding

Forward PCR primer

Reverse PCR primer

P5 sequence that will bind to the flow cell

Tag that allows for P5 and i5 to be incorporated by PCR and is the sequencing primer binding site

STR/SNP

Tag that allows for i7 and P7 to be incorporated by PCR and is the sequencing primer binding site

P7 sequence that will bind to the flow cell

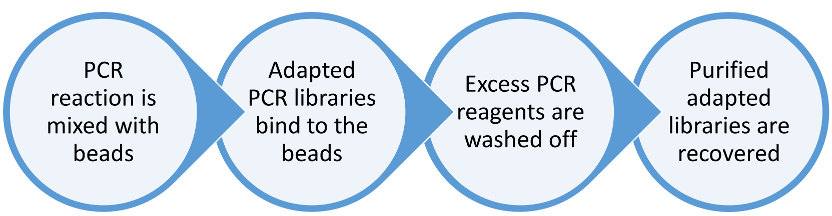
Workflow: ForenSeq™ DNA Signature Prep kit

Step: Purify libraries

Purpose: Purify amplified libraries

Prepare bead suspension
Pipette 45 µL of bead suspension into plate
Pipette 45 µL of PCR into bead
Shake 1800 rpm for 2 min
Let sit for 5 min
Place on magnetic stand for 2 min (until clear)
Wash with 200 µL 80% EtOH – 2 times
Add 52.2 µL of resuspension buffer to each well
Shake 1800 rpm for 2 min
Place on magnetic stand for 2 min (until clear)
Recover 50 µL in a fresh plate

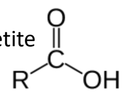
Result: purified adapted PCR products

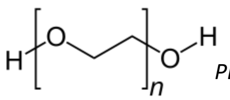


Beads! Beads! Beads!

Solid Phase Reversible Immobilization (SPRI)
Paramagnetic = magnetic only in a magnetic field

Polystyrene core surrounded by a layer of magnetite coated with carboxyl molecules



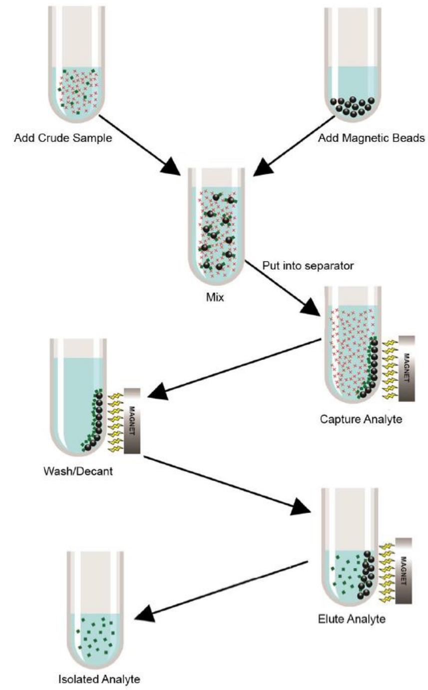


PEG

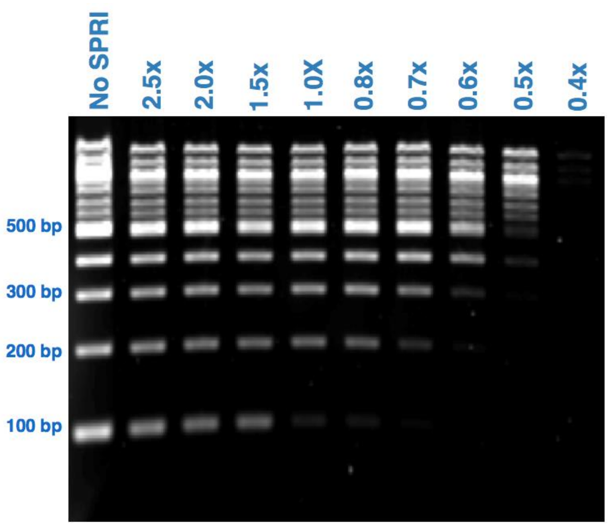
In the presence of PEG and salt (e.g. 20% PEG and 2.5 M NaCl) the DNA is driven to bind to the negatively charged surface

Wash or clean up (80% EtOH solution)
Resuspend

The ratio of SPRI:DNA can tune the length of DNA fragments bound



Sample purification and size selection can be accomplished using magnetic beads



As the concentration of beads **decreases** the size of the DNA captured **increases**

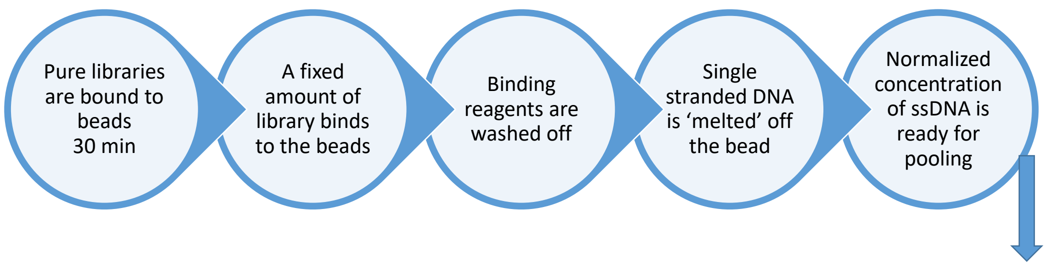
https://www.broadinstitute.org/genome-sequencing/broadillumina-genome-analyzer-boot-camp

Workflow: ForenSeq™ DNA Signature Prep kit

Step: Normalize libraries

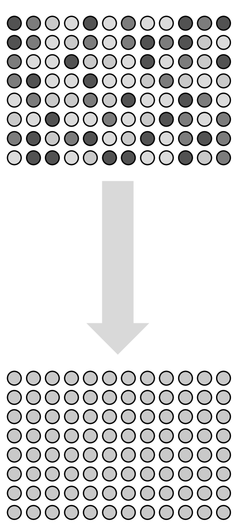
Purpose: Ensure that libraries of varying yields are equally represented within the sequencing run
By normalizing the concentration of the libraries, while preserving the content of each library, post-PCR quantification and individual PCR product normalization are not necessary

Mix beads and pure library – 30 min
Different than the previous purification – **one strand of the library adapter *may* be biotinylated (beads would be coated with streptavidin)**
Beads bind a fixed amount of library (ideally without bias; locus/adapter size)
2x EtOH wash
Wash with NaOH which denatures the library adapter – **leaving single stranded free in solution** (denatured and ready for pooling)
Recover and add to storage buffer



Approx. 2 nM of library post bead normalization

ForenSeq Library Preparation
Bead Normalization



Purified libraries:
Range of yields

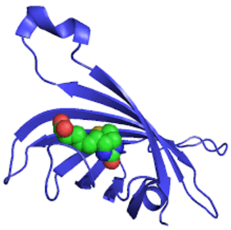
Bead-based Normalization
1. Equal volume of beads added to each well
2. Beads bind equal amount of product per well
3. Excess removed
4. Products eluted off beads

Normalized libraries:
Equally represented

Sample Pooling:
Pool 5 µl of each desired library

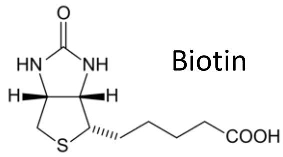
Slide courtesy of Meghan Didier (Verogen)

Streptavidin and Biotin



Streptavidin (SA)

+



Biotin

Dissociation constant (Kd) ≈ 10^{-14}

DNA can be labeled with biotin moieties
Magnetic beads can be coated with SA

The SA-Biotin complex is stable to organic solvents, denaturants, detergents, temperature, and pH

Very useful for biotech separation applications

Biotin could be attached at one of the 5' ends of the library molecule

https://en.wikipedia.org/wiki/Streptavidin
https://en.wikipedia.org/wiki/Biotin

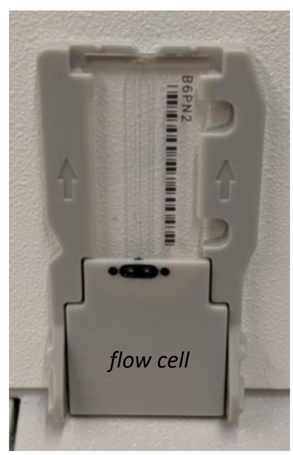
Workflow: ForenSeq™ DNA Signature Prep kit

Step: Pool libraries

Purpose: Combines equal volumes of normalized library to create a pool of libraries that are sequenced together on the same flow cell

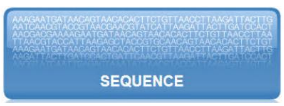
Collect and pool normalized libraries
Collect across the plate in a 8 strip tube – then pooled into a single tube

https://www.illumina.com/products/by-type/sequencing-kits.html



MiSeq FGx Sequencing

- Preloaded single use reagent cartridge
- Positive consumables tracking
- Auto flow cell positioning
- Walkaway automation



Slide courtesy of Meghan Didier (Verogen)

Workflow: ForenSeq™ DNA Signature Prep kit

Step: Sequencing

Purpose: Sequence the PCR products

Note: we will not be covering how to set up a specific instrument, loading the system, operational software, etc. This is covered in training materials, software 'wizards'

(Fluorescent) Sequencing by synthesis

- The library you created is hybridized to a flow cell
- Individual strands create 'clusters' through bridge amplification
- Sequencing proceeds <u>one base per cycle</u>
- Each A, G, C, T has a unique fluorescent dye attached
- Four images of the flow cell per cycle allows for the assignment of sequence at each cluster

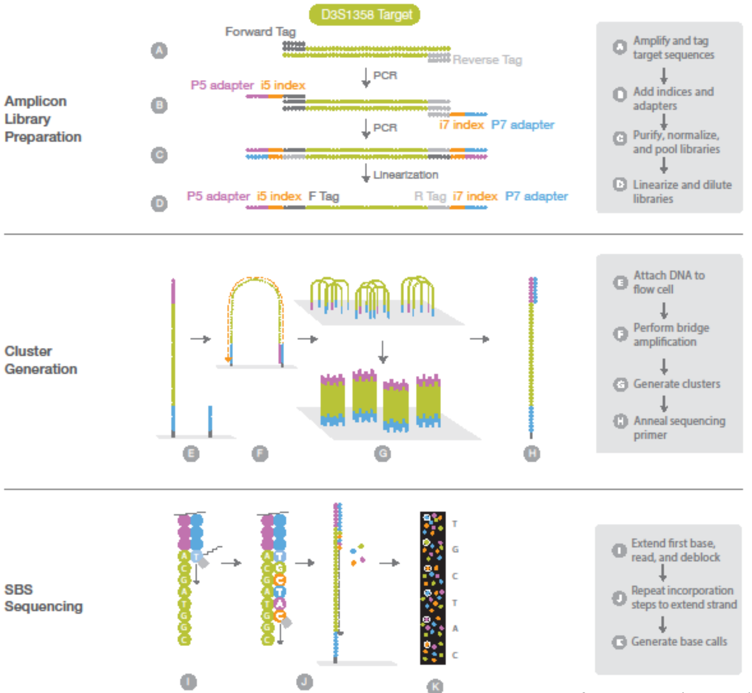


Figure courtesy of John Walsh (Verogen)

Library preparation

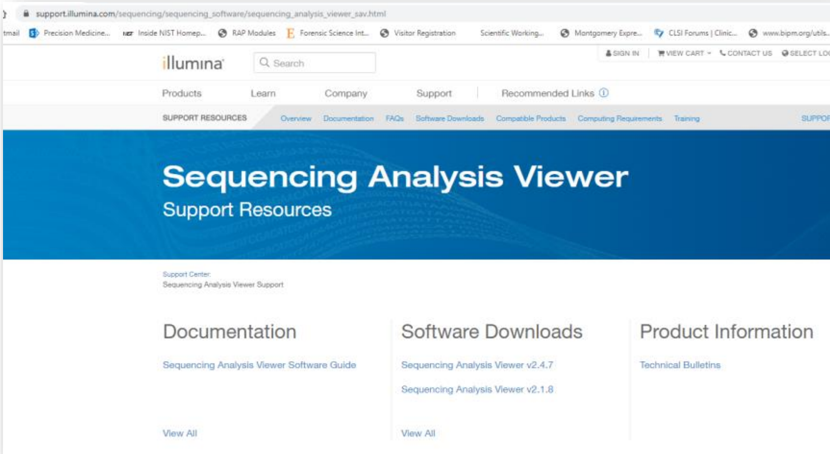
Creating a clusters that each represent a single library molecule

Sequencing by synthesis

Illumina Sequencing Movie

- https://www.youtube.com/watch?v=womKfikWlxM

Illumina tool for sequencing analysis (SAV)



View quality metrics

https://support.illumina.com/sequencing/sequencing_software/sequencing_analysis_viewer_sav.html

Data Analysis in the UAS

- Working with a cloud instance of UAS set up by Verogen (thank you!)
- Sequence data files from NIST SRM 2391d
- Intended to give a brief overview of the sequence data

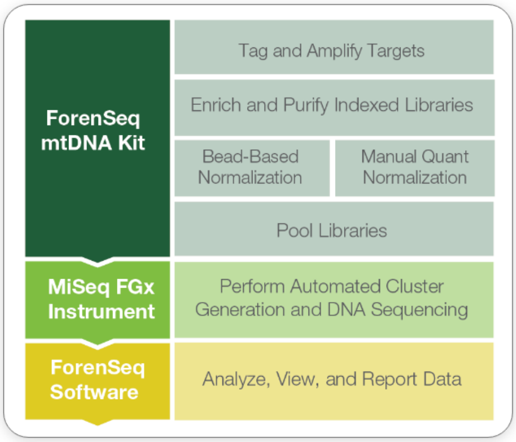
Summary of ancestry and phenotype marker estimates For NIST SRM 2391d

Component	ForenSeq			Precision ID				Mito	Y SNP
	Ancestry	*Hair*	*Eye*	*Ancestry*	*Hair*	*Hair*	*Eye*		
A	European	0.68	0.66	European	0.66	1.00 light	0.67	T2b3	-
B	African	0.69	0.86	African	0.66	0.93 light	0.85	L1c1a	E
C	African	0.84	1.00	African	0.68	1.00 dark	1.00	L1b1a	E
E	European	0.61	0.71	European/SW Asian	0.69	0.72 light	0.72	T2a3	-

Predictions made using vendor tools

Mito Sequencing

ForenSeq™ mtDNA Control Region Solution - MiSeq FGx

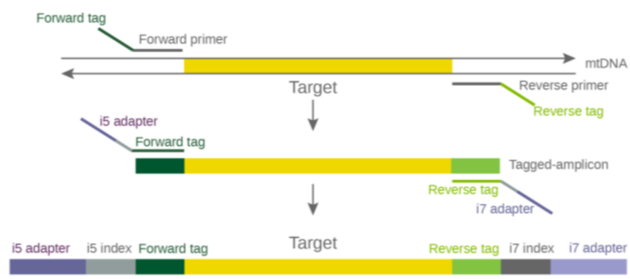


CR positions 16024 - 576

PCR products < 150 bp (average 118 bp)
18 primary amplicons; 122 primers

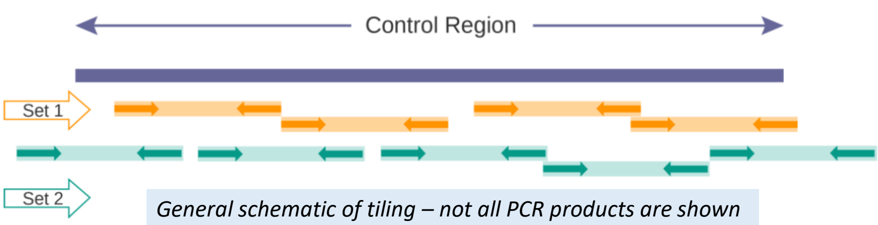
Two unique targeted PCR reactions are set up per sample

Figure 1 ForenSeq mtDNA Control Region Prep Overview



Tiling across the Control Region

Figure 2 Schematic of Tiled Amplicons for Complete Coverage Across the Control Region



Smaller PCR product sizes help with degraded samples
Tiling and overlapping PCR products to cover regions under primers
Degenerate primer sets allow PCR primers to bind to different 'mito types'

Degenerate PCR primers

Primer A

Individual #1

```
                                               CTACGATCGACTAGCATCGAC
TGTAGCTGATGCAGTCAGTGCTAGCTGATGCGTCAGTCGATGCTAGCTGATCGTAGCTGATGCT
```

Primer B

Individual #2

```
                                               CTACGATCGCCTAGCATCGAC
TGTAGCTGATGCAGTCAGTGCTAGCTGATGCGTCAGTCGATGCTAGCGGATCGTAGCTGATGCT
```

Degenerate primers in a PCR primer mix will allow for the efficient amplification across polymorphic regions in a genome.
Reduces the chance of dropout

Two PCR set up

Figure 4: ForenSeq Sample Plate Setup for 48 samples

	1	2	3	4	5	6	7	8	9	10	11	12
	Set 1 Primer Mix						Set 2 Primer Mix					
	R713	R714	R716	R717	R718	R719	R713	R714	R716	R717	R718	R719
A501	Sample 1	Sample 9	Sample 17	Sample 25	Sample 33	Sample 41	Sample 1	Sample 9	Sample 17	Sample 25	Sample 33	Sample 41
A502	Sample 2	Sample 10	Sample 18	Sample 26	Sample 34	Sample 42	Sample 2	Sample 10	Sample 18	Sample 26	Sample 34	Sample 42
A503	Sample 3	Sample 11	Sample 19	Sample 27	Sample 35	Sample 43	Sample 3	Sample 11	Sample 19	Sample 27	Sample 35	Sample 43
A504	Sample 4	Sample 12	Sample 20	Sample 28	Sample 36	Sample 44	Sample 4	Sample 12	Sample 20	Sample 28	Sample 36	Sample 44
A505	Sample 5	Sample 13	Sample 21	Sample 29	Sample 37	Sample 45	Sample 5	Sample 13	Sample 21	Sample 29	Sample 37	Sample 45
A506	Sample 6	Sample 14	Sample 22	Sample 30	Sample 38	Sample 46	Sample 6	Sample 14	Sample 22	Sample 30	Sample 38	Sample 46
A507	Sample 7	Sample 15	Sample 23	Sample 31	Sample 39	Sample 47	Sample 7	Sample 15	Sample 23	Sample 31	Sample 39	Sample 47
A508	Sample 8	Sample 16	Sample 24	Sample 32	Sample 40	Sample 48	Sample 8	Sample 16	Sample 24	Sample 32	Sample 40	Sample 48

Two separate PCRs

An individual sample will have the same i5/i7 index for both PCRs

Summary

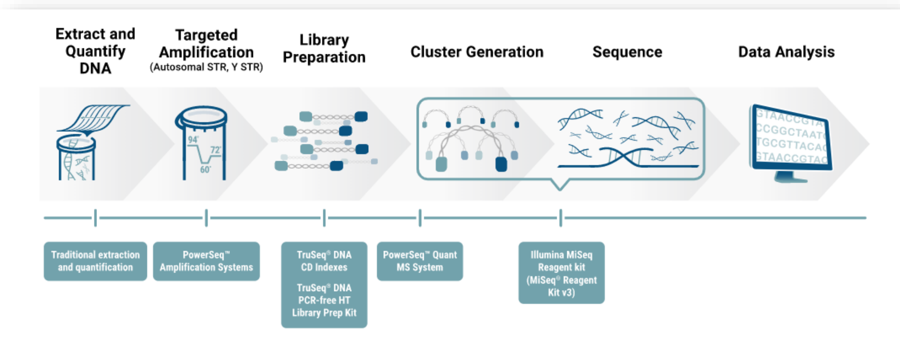
- Uses PCR tagging to incorporate library adapters
- Bead-based normalization is used to ensure that each library is of a similar concertation prior to sequencing
- Care should be taken with PCR 2 and i5/i7 indices to avoid contamination
- Sequencing by synthesis – all four bases are incorporated and read per cycle

NGS workflow #2

	assay	instrument	software	markers
Promega/Illumina	PowerSeq 46 GY System	MiSeq	STRait Razor toaSTR	STRs

Also purchase
Illumina *TruSeq DNA PCR-Free HT Sample Preparation Kit*
or equivalent from another vendor

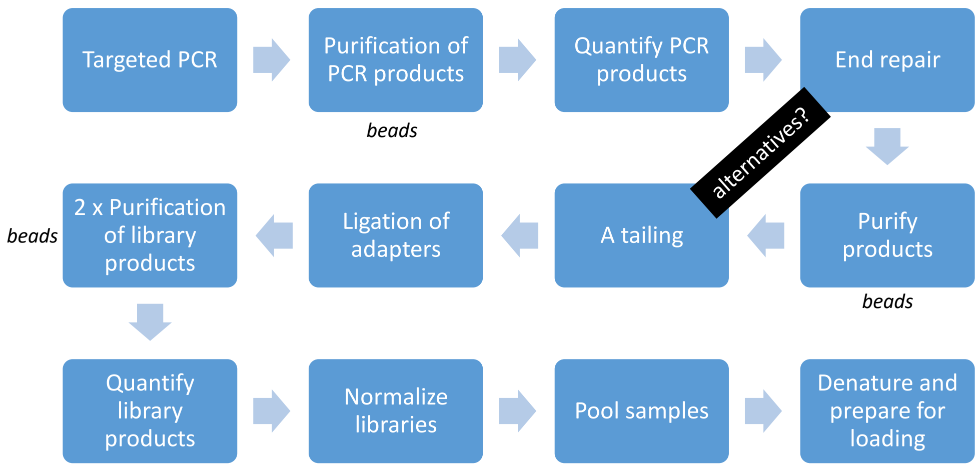
PowerSeq^{TM} 46GY System



Locus Name	Locus Name
D8S1179	DYS391
D21S11	DYS19
D7S820	DYS385ab
CSF1PO	DYS389I/II
D3S1358	DYS390
TH01	DYS392
D13S317	DYS393
D16S539	DYS437
D2S1338	DYS438
D19S433	DYS439
vWA	DYS448
TPOX	DYS456
D18S51	DYS458
D5S818	DYS481
FGA	DYS533
Penta D	DYS549
Penta E	DYS570
Amelogenin	DYS576
D1S1656	DYS635
D2S441	DYS643
D10S1248	Y-GATA-H4
D12S391	
D22S1045	
DYS391	

PCR products for each locus are designed to be in a range of 140–300bp

PowerSeq 46GY (Sequenced on MiSeq FGx)



Workflow: Promega PowerSeq 46GY

Step: Targeted PCR amplification

Purpose: Targeted amplification of STRs

Target 0.5 ng of genomic DNA
PowerSeq PCR primers and mastermix
25 µL volume
30 cycle PCR

I don’t believe there are any special tails or adapters on the PCR primers

Thermal Cycling Protocol

96°C for 1 minute, then:

30 cycles of
96°C for 5 seconds,
60°C for 35 seconds,
72°C for 5 seconds, then:

60°C for 2 minutes

4°C soak

Workflow: Promega PowerSeq 46GY

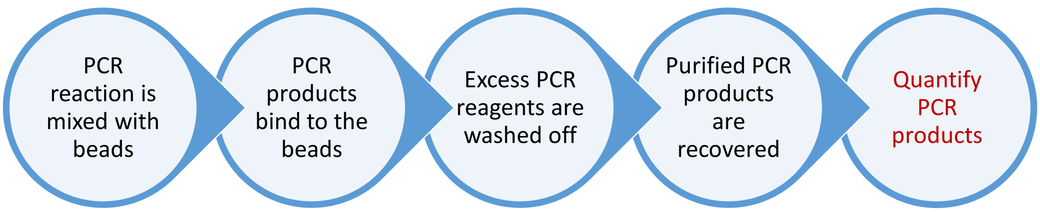
Step: PCR product purification

Purpose: Remove excess PCR primers and buffers

Sample purification beads from the:
Illumina *TruSeq DNA PCR-Free HT Sample Preparation Kit*
Proteinase K is added to avoid bead clumping

From the manual: *Optimal yield and balanced representation of all STR loci is dependent on thorough mixing of the beads and samples*

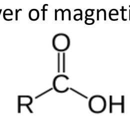
Add purification bead solution to sample (vortex to keep homogeneous)
Mix (pipette up and down 10 times)
Place tube(s) on magnetic stand/separator for 5 minutes
Discard solution
Wash 2 times with 80% EtOH (don't disturb the pellet)
Resuspend in buffer (30 µL) (don't disturb the pellet)



Beads! Beads! Beads!

Solid Phase Reversible Immobilization (SPRI)
Paramagnetic = magnetic only in a magnetic field

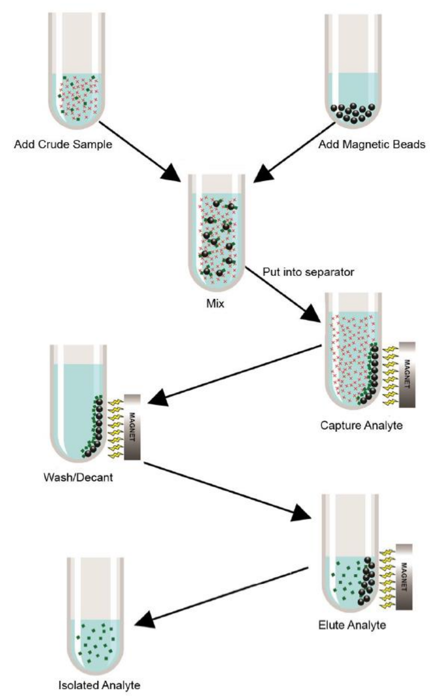
Polystyrene core surrounded by a layer of magnetite
Coated with carboxyl molecules



In the presence of PEG and salt (e.g. 20% PEG and 2.5 M NaCl) the DNA is driven to bind to the negatively charged surface

Wash or clean up (80% EtOH solution)
Resuspend

The ratio of SPRI:DNA can tune the length of DNA fragments bound



Workflow: Promega PowerSeq 46GY

Step: Quantify PCR products

Purpose: Estimate PCR product concentration for library preparation

From the manual: *Determine the concentration of the purified amplification products by measuring the absorbance using a fluorescence-based quantification method. We recommend the Quantifluor® dsDNA System (Promega Cat.# E2670) or the QuantiFluor® One dsDNA System (Promega Cat.# E4871 or E4870) with the Quantus™ Fluorometer or the GloMax® Multi Detection System].*

A fluorometer can be used to obtain a concertation estimate of purified PCR products

Single measurements or plate format

An aliquot of the PCR products is mixed with a dsDNA binding dye and the fluorescence is measured

Based on the measurement, the PCR products *for each sample* are diluted to:

Protocol target: 500 ng in a volume of 60 µL

Quantification (fluorometer)

SpectraMax M2 Microplate reader
Molecular Devices

Estimate [DNA]
Dyes that emit fluorescence upon binding dsDNA

Qubit
Thermo Fisher

Quantus
Promega

GloMax
Promega

Workflow: Promega PowerSeq 46GY

Step: End Repair

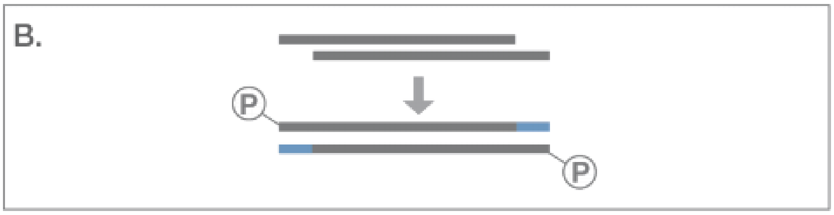
Purpose: Ensure that products are free of overhangs and phosphorylation of 5' ends

A cocktail of enzymes (general)
An optimized mixture of:

- Klenow Fragment (fill in 5' – 3' overhangs)
- T4 DNA Polymerase (remove 3' overhangs)
- T4 Polynucleotide Kinase (phosphorylation of DNA ends)
- Reaction buffer containing ATP and dNTPs

Component	Volume per Sample
500ng purified amplification product diluted in Resuspension Buffer (from step 4)	60µl
End Repair Mix 2	40µl
Total Volume	100µl

30 min incubation @30ºC



Blunt-end fragments are created, 5' ends are phosphorylated
Now ready for 3' A-tailing…**post bead clean up**

Workflow: Promega PowerSeq 46GY

Step: A-tailing (post bead cleanup of end repair)

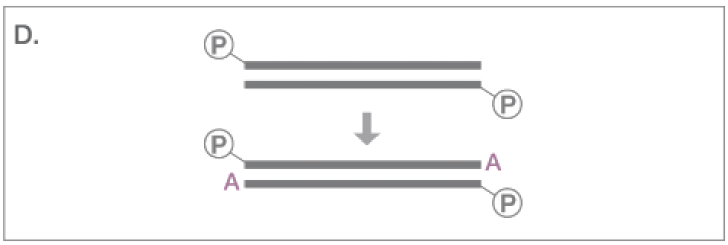
Purpose: Adding a non-templated nucleotide to the 3' end of a blunt, double-stranded DNA molecule

An optimized mixture of:

- Taq DNA Polymerase
- dATP

Component	Volume per Sample
Purified end-repair reaction (from step 9.17)	17.5µl
A-Tailing Mix	12.5µl
Total Volume	30µl

37°C for 30 minutes
70°C for 5 minutes



'A' base is added
Preparing the PCR products ready for ligation and to reduce the chance of concatemers forming

Note: there are some kits that combine End Repair and A-tailing and reduce some of the bead clean steps

KAPA HyperPlus Kit (now owned by Roche)
KAPA HyperPrep Kit (now owned by Roche)

General: commercial products are evolving to reduce the steps/time in sequencing library preparation

https://sequencing.roche.com/en-us/products-solutions/by-category/library-preparation/dna-library-preparation/kapa-hyperplus.html

Workflow: Promega PowerSeq 46GY

Step: Ligation of adapters

Purpose: Adding the adapters containing P5/P7, indices, and sequencing primer regions

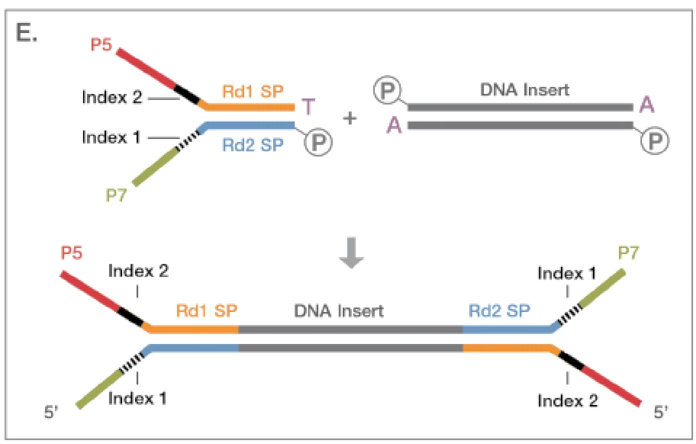
An optimized mixture of:

- DNA Ligase, buffers, ATP
- Dual-index adapters – these are in a plate
- Each well has a unique index set (i5/i7)

- 30°C for 20 minutes

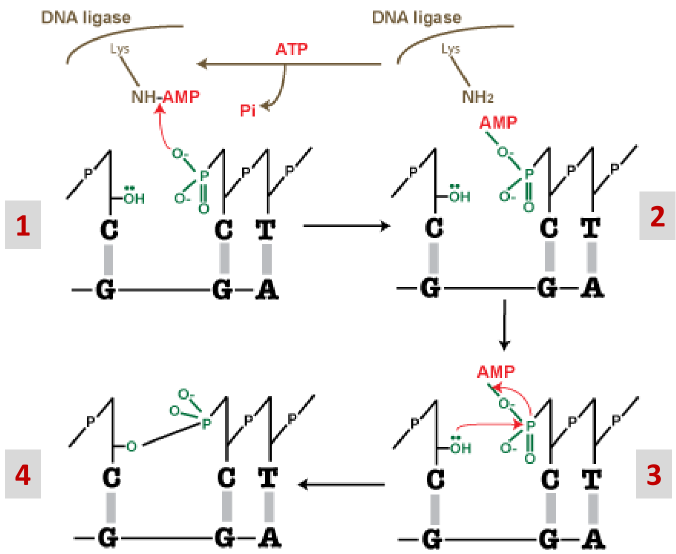
- One more bead clean up

- **And purify again!**



Dual-index adapters are ligated to the fragments* and final product is ready for cluster generation.

https://www.illumina.com/documents/products/datasheets/datasheet_truseq_dna_pcr_free_sample_prep.pdf



DNA Ligation

DNA ligase catalyzes the joining of the 3′-OH to the 5′-phosphate via a two step mechanism

First, the AMP nucleotide, which is attached to a lysine residue in the enzyme's active site, is transferred to the 5′-phosphate

Then the AMP-phosphate bond is attacked by the 3′-OH, forming the covalent bond and releasing AMP

To allow the enzyme to carry out further reactions the AMP in the enzyme's active site must be replenished by ATP.

Workflow: Promega PowerSeq 46GY

Step: Quantification of the library products

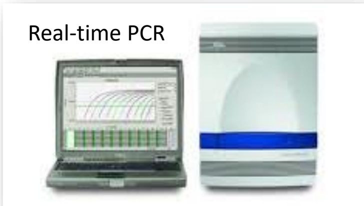
Purpose: Determine the concentration of each library (prior to normalization)

Quantification (qPCR-based)

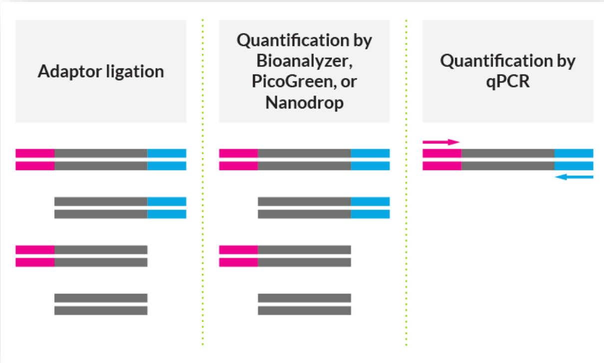
PowerSeq™ Quant MS System

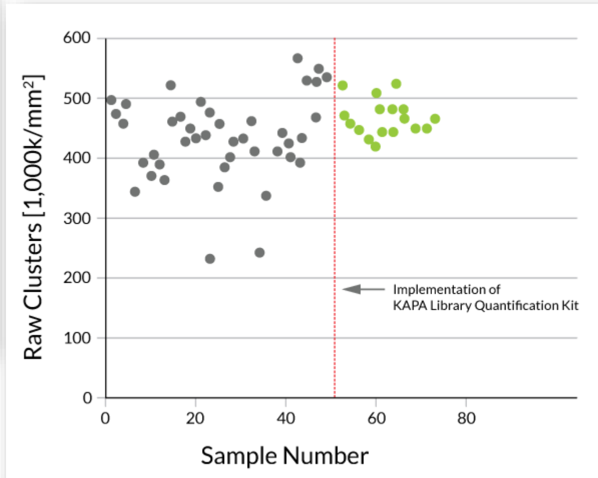
- Quantification of MiSeq® Platform Compatible MPS Libraries enables normalization of MPS libraries based on DNA quantification
- Uses BRYT Green® dye-based qPCR system for maximum sensitivity and reproducibility
- Enables accurate and balanced multiplexed Illumina pooled libraries

KAPA Library Quantification Kits
for Next-Generation Sequencing



Quantification (qPCR-based)





https://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/library-quantification/

Workflow: Promega PowerSeq 46GY

Step: Normalization of libraries, pooling, denature and load

Purpose: Normalize the concentration of each library (followed by pooling)

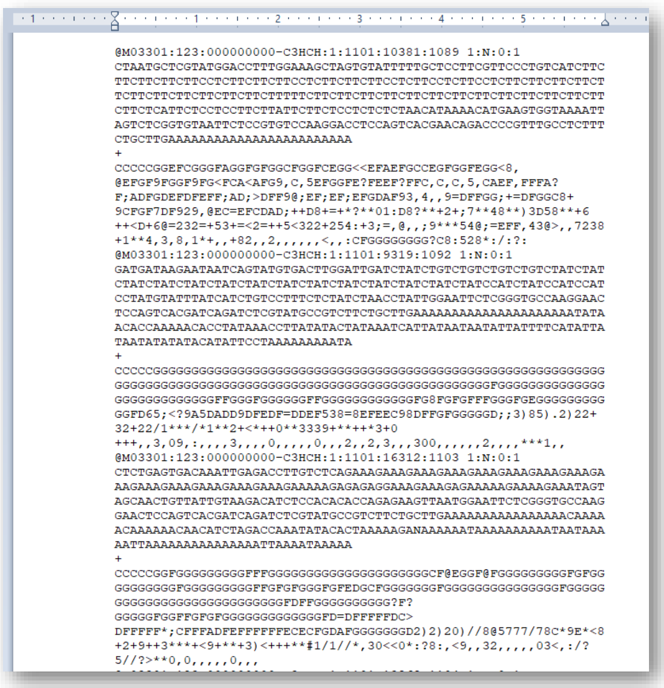
Based on the DNA quantification results obtained, normalize the DNA concentration for each library to 4nM with 10mM Tris-HCl (pH 8.5)

Pool equal volumes of the 4nM libraries

Pooled libraries are denatured with 0.2 N NaOH and prepared for loading onto the MiSeq cartridge

Already covered Illumina sequencing…

What is a FASTQ file?



```
@M03301:123:000000000-C3HCH:1:1101:10381:1089 1:N:0:1
CTAATGCTCGTATGGACCTTTGGAAAGCTAGTGTATTTTTGCTCCTTCGTTCCCTGTCATCTTC
TTCTTCTTCTTCCTCTTCTTCTTCCTCTTCTTCTTCCTCTTCCTCTTCCTCTTCTTCTTCTTCT
TCTTCTTCTTCTTCTTCTTCTTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
CTTCTCATTCTCCTCCTTCTTATTCTTCTCCTCTCTCTAACATAAAACATGAAGTGGTAAAATT
AGTCTCGGTGTAATTCTCCGTGTCCAAGGACCTCCAGTCACGAACAGACCCCGTTTGCCTCTTT
CTGCTTGAAAAAAAAAAAAAAAAAAAAAAAA
+
CCCCCGGEFCGGGFAGGFGFGGCFGGFCEGG<<EFAEFGCCEGFGGFEGG<8,
@EFGF9FGGF9FG<FCA<AFG9,C,5EFGGFE?FEEF?FFC,C,C,5,CAEF,FFFA?
F;ADFGDEFDFEFF;AD;>DFF9@;EF;EF;EFGDAF93,4,,9=DFFGG;+=DFGGC8+
9CFGF7DF929,@EC=EFCDAD;++D8+=+*?**01:D8?**+2+;7**48**)3D58**+6
++<D+6@=232=+53+=<2=++5<322+254:+3;=,@,,;9***54@;=EFF,43@>,,7238
+1**4,3,8,1*+,,+82,,2,,,,,,<,,:CFGGGGGGG?C8:528*:/:?:
@M03301:123:000000000-C3HCH:1:1101:9319:1092 1:N:0:1
GATGATAAGAATAATCAGTATGTGACTTGGATTGATCTATCTGTCTGTCTGTCTGTCTATCTAT
CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCCATCTATCCATCCAT
CCTATGTATTTATCATCTGTCCTTTCTCTATCTAACCTATTGGAATTCTCGGGTGCCAAGGAAC
TCCAGTCACGATCAGATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAAAAAAAATATA
ACACCAAAAACACCTATAAACCTTATATACTATAAATCATTATAATAATATTATTTTCATATTA
TAATATATATACATATTCCTAAAAAAAAAATA
+
CCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGFGGGGGGGGGGGGGG
GGGGGGGGGGGGGFFGGGFGGGGGGFFGGGGGGGGGGGGGFG8FGFGFFFGGGFGEGGGGGGGGG
GGFD65;<?9A5DADD9DFEDF=DDEF538=8EFEEC98DFFGFGGGGGD;;3)85).2)22+
32+22/1***/*1**2+<*++0**3339+****+*3+0
+++,,3,09,:,,,,3,,,,0,,,,,0,,,2,,2,3,,,300,,,,,,2,,,,***1,,
@M03301:123:000000000-C3HCH:1:1101:16312:1103 1:N:0:1
CTCTGAGTGACAAATTGAGACCTTGTCTCAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA
AAGAAAGAAAGAAAGAAAGAAAGAAAAAGAGAGAGGAAAGAAAGAGAAAAAGAAAAGAAATAGT
AGCAACTGTTATTGTAAGACATCTCCACACACCAGAGAAGTTAATGGAATTCTCGGGTGCCAAG
GAACTCCAGTCACGATCAGATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAAACAAAA
ACAAAAAACAACATCTAGACCAAATATACACTAAAAAGANAAAAAATAAAAAAAAAAATAATAAA
AATTAAAAAAAAAAAAAAAATTAAAATAAAAA
+
CCCCCGGFGGGGGGGGGFFFGGGGGGGGGGGGGGGGGGGGGCF@EGGF@FGGGGGGGGGFGFGG
GGGGGGGGFGGGGGGGGGFFGFGFGGGFGFEDGCFGGGGGGGFGGGGGGGGGGGGGGGFGGGGG
GGGGGGGGGGGGGGGGGGGGGFDFFGGGGGGGGGG?F?
GGGGGFGGFFGFGFGGGGGGGGGGGGGFD=DFFFFFDC>
DFFFFF*;CFFFADFEFFFFFFECECFGDAFGGGGGGGD2)2)20)//8@5777/78C*9E*<8
+2+9++3***+<9+***+3)<++++**#1/1//*,30<<0*:?8:,<9,,32,,,,,03<,:/?
5//?>**0,0,,,,,0,,,
```

FASTQ Format

- FASTQ - normally uses four lines per sequence.

- Line 1 begins with a '@' character and is followed by a sequence identifier and an *optional* description
- Line 2 is the raw sequence letters.
- Line 3 begins with a '+' character and is *optionally* followed by the same sequence identifier (and any description) again.
- Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence.

- ```
  (1)@SEQ_ID
  (2)GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCC
  (3)+
  (4)!''*((((***+))%%%++)(%%%%).1***-+*''))**
  ```

http://maq.sourceforge.net/fastq.shtml

Data analysis using open source software/tools
  ```

Tools that focus and STR/SNPs

Name	Availability
Freeware	
FDSTools	Python Package Index
Seqmapper	http://forensic.mc.ntu.edu.tw:9000/SEQMapperWeb/Default.aspx
STRait Razor v2s STRait Razor 3.0	https://www.unthsc.edu/graduate-school-of-biomedical-sciences/laboratory-faculty-and-staff/strait-razor/
STRinNGS	Upon request from the University of Copenhagen
toaSTR	https://www.toastr.de
For purchase	
ExactID	https://www.battelle.org/government-offerings/homeland-security-public-safety/security-law-enforcement/forensic-genomics/exactid
GeneMarkerHTS	https://softgenetics.com/GeneMarkerHTS.php
Armed Expert Mixture Ace	https://nichevision.com/mixtureace/
Assay specific, for purchase	
Converge	https://www.thermofisher.com/order/catalog/product/A35131
Universal Analysis Software	https://verogen.com/products/

Table courtesy of Katherine Gettings (NIST)

Name	References
Freeware	
FDSTools	*J. Hoogenboom, K.J. van der Gaag, R.H. de Leeuw, T. Sijen, P. de Knijff, J.F. Laros, FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise, Forensic Sci Int Genet 27 (2017) 27-40.*
Seqmapper	*J.C. Lee, B. Tseng, L.K. Chang, A. Linacre, SEQ Mapper: A DNA sequence searching tool for massively parallel sequencing data, Forensic Sci Int Genet 26 (2017) 66-69.*
STRait Razor v2s STRait Razor 3.0	*J.L. King, F.R. Wendt, J. Sun, B. Budowle, STRait Razor v2s: Advancing sequence-based STR allele reporting and beyond to other marker systems, Forensic Sci Int Genet 29 (2017) 21-28.* *A.E. Woerner, J.L. King, B. Budowle, Fast STR allele identification with STRait Razor 3.0, Forensic Science International: Genetics (2017).*
STRinNGS	*S.L. Friis, A. Buchard, E. Rockenbauer, C. Borsting, N. Morling, Introduction of the Python script STRinNGS for analysis of STR regions in FASTQ or BAM files and expansion of the Danish STR sequence database to 11 STRs, Forensic Sci Int Genet 21 (2016) 68-75.*
toaSTR	*S. Ganschow, J. Silvery, J. Kalinowski, C. Tiemann, toaSTR: A web application for forensic STR genotyping by massively parallel sequencing, Forensic Sci Int Genet 37 (2018) 21-28.*

Table courtesy of Katherine Gettings (NIST)

STRait Razor 3.0


Forensic Science International: Genetics 30 (2017) 18–23

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

Fast STR allele identification with STRait Razor 3.0

August E. Woerner[a,*], Jonathan L. King[a], Bruce Budowle[a,b]

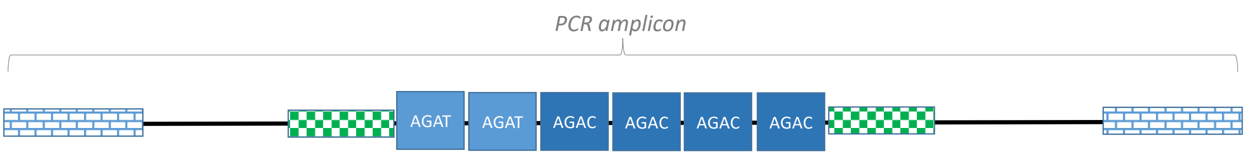
[a] Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[b] Center of Excellence in Genomic Medicine (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

- https://github.com/Ahhgust/STRaitRazor
- Code to run STRait Razor 3.0

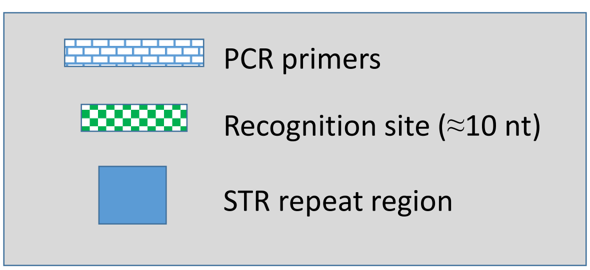
- Links an excel file to parse the results
- https://www.dropbox.com/s/t3n0d2h6od0qek2/STRait%20Razor%20Analysis%20v3.xlsm?dl=1

A little bit of informatics

Recognition Site-Based Informatics for STRs

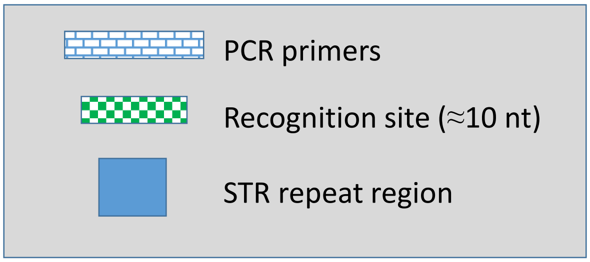


Software returns:
The length between the recognition sequences (= 24)
A reference table returns a "6" allele **and the sequence between the recognition sites**



Recognition Site-Based Informatics for STRs

Moving the recognition sites out further:
Captures the flanking region SNPs and Indels
Still returns the allele length and **more sequence between the recognition sites**



STRait Razor file example-TPOX

1. Locus : allele call
2. Length of sequence
3. Sequence
4. Coverage

TPOX:7	28 bases	AATGAATGAATGAATGAATGAATGAATG	9
TPOX:7	28 bases	AATGAATGAATGAGTGAATGAATGAATG	1
TPOX:8	32 bases	AATGAATGAATGAATGAATGAATGAATGAATG	332
TPOX:8	32 bases	AATGAATGAATGAATGAATGAATGAATGAAAG	2
TPOX:8	32 bases	AATGAATGAATGAATGAATGAAGGAATGAATG	1
TPOX:8	32 bases	AATGAATGAATGAATGACTGAATGAATGAATG	1
TPOX:8	32 bases	AACGAATGAATGAATGAATGAATGAATGAATG	1
TPOX:8	32 bases	AATGAATGAATGAATGAGTGAATGACTGATTG	1
TPOX:8	32 bases	AATGAATGAATGAATGAATGAATGAATGGATG	1
TPOX:8	32 bases	AATGAATGATTGAATGAATGAATGAATGAATG	1
TPOX:8	32 bases	AATGAATGAATGAATGAATGAATGAATGTATT	1
TPOX:8	32 bases	AATGAATGAATGGATGAATGAATGAATGAATG	1
TPOX:8	32 bases	AATGAATGAACGAATGACTGAATGAATGAATG	1
TPOX:9	36 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATG	17
TPOX:10	40 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATGAATG	321
TPOX:10	40 bases	ACTGAATGAATGAATGAATGAATGAATGAATGAATGAATG	1
TPOX:10	40 bases	AATGAATGAATGAATGAATGAATGAATGAAAGAATGAATG	1
TPOX:10	40 bases	AATGAATGAATGAGTGAATGAATGAATGAATGAATGAATG	1
TPOX:10	40 bases	AATGAATGAATAAATGAATGAATGAATGAATGAATGAATG	1
TPOX:10	40 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATGATTG	1
TPOX:10	40 bases	AATGAATGAATGCATGAATGAATGAATGAATGAATGAATG	1
TPOX:11	44 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG	2

STRait Razor file example-TPOX (8,10)

1. Locus : allele call
2. Length of sequence
3. Sequence
4. Coverage

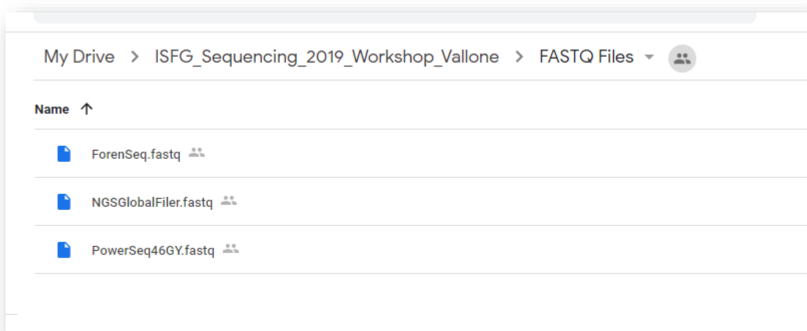
Locus	Length	Sequence	Coverage	Comment
TPOX:7	28 bases	AATGAATGAATGAATGAATGAATGAATG	9	n-4 stutter
TPOX:7	28 bases	AATGAATGAATGA**G**TGAATGAATGAATG	1	
TPOX:8	32 bases	AATGAATGAATGAATGAATGAATGAATGAATG	332	Allele
TPOX:8	32 bases	AATGAATGAATGAATGAATGAATGAATGAA**A**G	2	
TPOX:8	32 bases	AATGAATGAATGAATGAATGAA**G**GAATGAATG	1	
TPOX:8	32 bases	AATGAATGAATGAATGA**C**TGAATGAATGAATG	1	
TPOX:8	32 bases	AA**C**GAATGAATGAATGAATGAATGAATGAATG	1	
TPOX:8	32 bases	AATGAATGAATGAATGA**G**TGAATGACTGA**T**TG	1	
TPOX:8	32 bases	AATGAATGAATGAATGAATGAATGAATG**G**ATG	1	
TPOX:8	32 bases	AATGAATGA**T**TGAATGAATGAATGAATGAATG	1	
TPOX:8	32 bases	AATGAATGAATGAATGAATGAATGAATG**T**AT**T**	1	
TPOX:8	32 bases	AATGAATGAATG**G**ATGAATGAATGAATGAATG	1	
TPOX:8	32 bases	AATGAATGAA**C**GAATGA**C**TGAATGAATGAATG	1	
TPOX:9	36 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATG	17	n-4 stutter; n+4 sutter
TPOX:10	40 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATGAATG	321	Allele
TPOX:10	40 bases	A**C**TGAATGAATGAATGAATGAATGAATGAATGAATGAATG	1	
TPOX:10	40 bases	AATGAATGAATGAATGAATGAATGAATGAA**A**GAATGAATG	1	
TPOX:10	40 bases	AATGAATGAATGA**G**TGAATGAATGAATGAATGAATGAATG	1	
TPOX:10	40 bases	AATGAATGAAT**A**AATGAATGAATGAATGAATGAATGAATG	1	
TPOX:10	40 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATGA**T**TG	1	
TPOX:10	40 bases	AATGAATGAATG**C**ATGAATGAATGAATGAATGAATGAATG	1	
TPOX:11	44 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATG	2	n+4 stutter?

Google Drive

- https://drive.google.com/drive/folders/1cbqsnBaEgivRkYIxNr2r2LtBFibvafvk

Or

- https://tinyurl.com/yyfwe8jc



toaSTR [https://www.toastr.de/]

ELSEVIER

Forensic Science International: Genetics
Volume 37, November 2018, Pages 21-28

toaSTR: A web application for forensic STR genotyping by massively parallel sequencing

Sebastian Ganschow [a], Janine Silvery [a], Jörn Kalinowski [b], Carsten Tiemann [a]

⊞ Show more

https://doi.org/10.1016/j.fsigen.2018.07.006

Get rights and content

Analyze PowerSeq Data

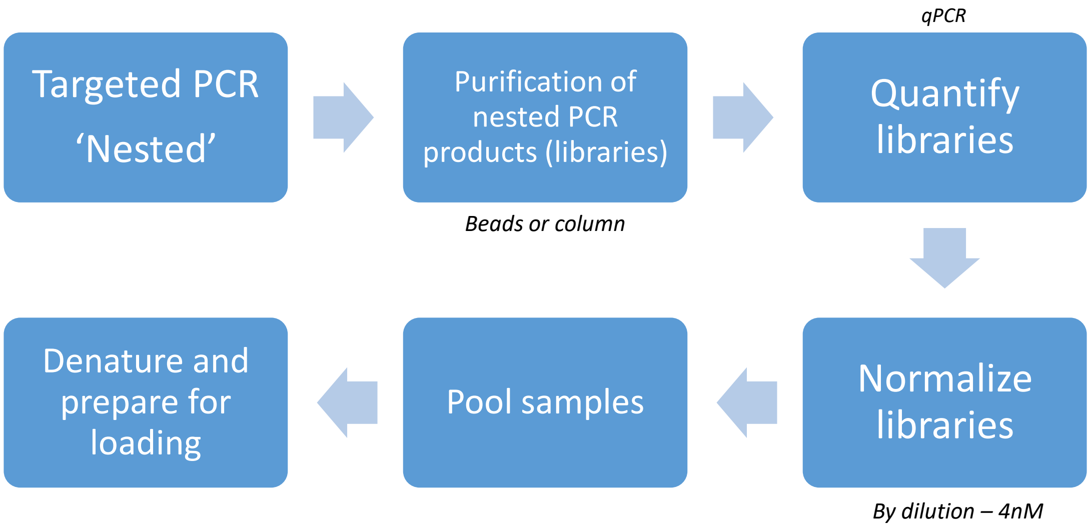
- STRait Razor 3.0
  - Run code
  - Excel file - view
- toaSTR
  - Upload
  - View results
  - Export

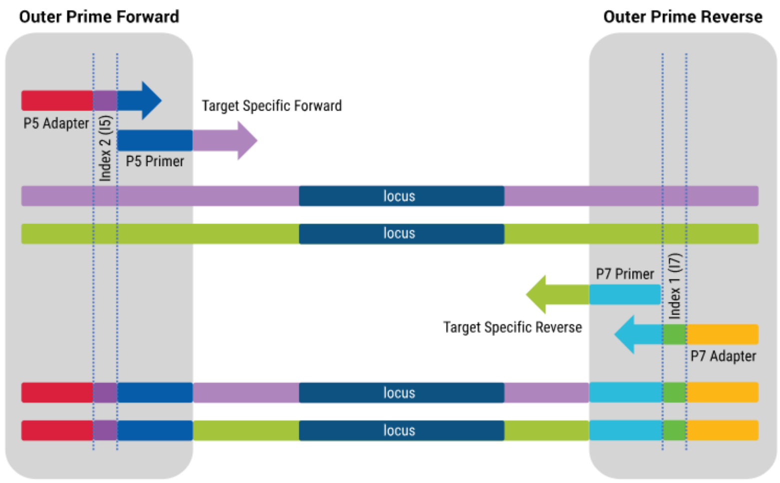
PowerSeq™ CRM Nested System, Custom

- Sequencing of mitochondrial control region
- *Nested* amplification protocol greatly reduces number of steps and time required for library preparation
  - Single step PCR
- 10 PCR products ranging from 144–237 bp

PowerSeq CRM Nested System

(Sequenced on MiSeq FGx)





Adapted from Lange *et al.* (2014) *BMC Genomics* **15**, 63.

Targeted PCR – one step

Index primers are provided in the kit

In preparing the PCR:
Mastermix + 10plex PCR primers

Add index primer 1 (i7 index)

Add index primer 2 (i5 index)

Then add template DNA

30 cycles of PCR

https://www.promega.com/-/media/files/resources/protocols/custom-systems/an322-powerseq-crm-system-custom.pdf

- Indexing Primers
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D701
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D702
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D703
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D704
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D705
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D706
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D707
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D708
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D709
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D710
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D711
  - 1 × 75µl PowerSeq™ Nested System Index Primer 1 D712
  - 1 × 75µl PowerSeq™ Nested System Index Primer 2 D501
  - 1 × 75µl PowerSeq™ Nested System Index Primer 2 D502
  - 1 × 75µl PowerSeq™ Nested System Index Primer 2 D503
  - 1 × 75µl PowerSeq™ Nested System Index Primer 2 D504
  - 1 × 75µl PowerSeq™ Nested System Index Primer 2 D505
  - 1 × 75µl PowerSeq™ Nested System Index Primer 2 D506
  - 1 × 75µl PowerSeq™ Nested System Index Primer 2 D507
  - 1 × 75µl PowerSeq™ Nested System Index Primer 2 D508

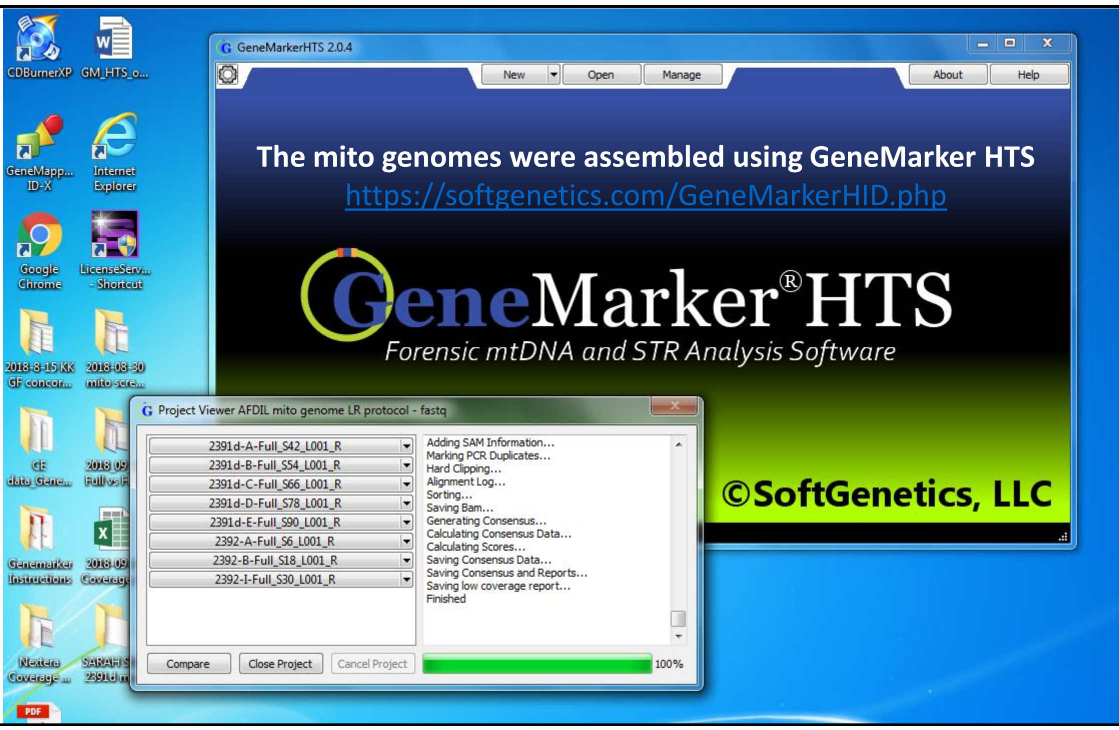
Adapter/Index primers are included in the kit

From the Promega manual

View Mito sequence data in IGV

- https://igv.org/app/

- Or download desktop version
- http://software.broadinstitute.org/software/igv/



Summary

- Uses ligation to incorporate library adapters
  - End repair and A-tailing
- Quantification steps for PCR products and libraries
  - Fluorescence readout
  - qPCR
- Both ‘top’ and ‘bottom’ strands are sequenced

NGS workflow #3

	assay	instrument	software	markers
Thermo Fisher	Precision ID Panels	S5	Converge	STRs, SNPs

Precision ID Panels – Thermo Fisher

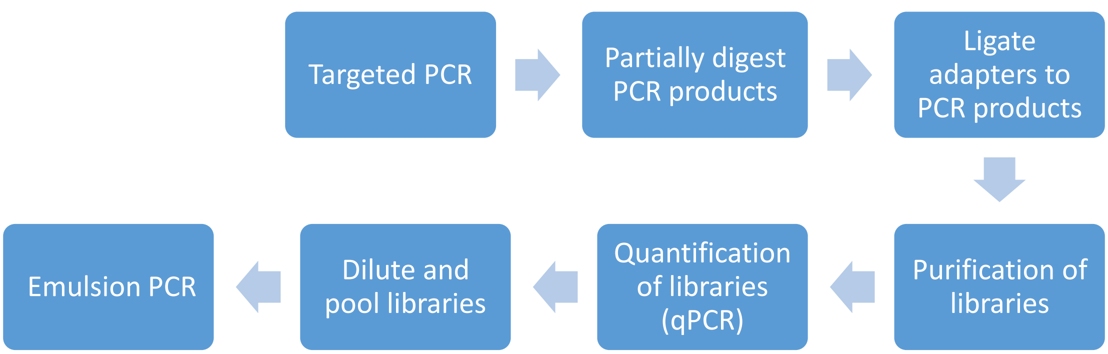
- STRs
  - GlobalFiler NGS STR Panel v2 (35 markers)
- SNPs
  - Identity (90 + 34 Y CHR), Ancestry (165), & Phenotype Panels (24)
- Mitochondrial Genome
  - Control Region Panel
  - Whole Genome Panel

Ion S5 System

Precision ID Panels (Sequenced on S5)

The clonal PCR does not take place on the sequencer (as compared to the MiSeq)

Precision ID Panels (Sequenced on S5)



Precision ID Panels (Sequenced on S5)

Targeted PCR → Partially digest PCR products → Ligate adapters to PCR products → Purification of libraries → Quantification of libraries (qPCR) → Dilute and pool libraries → Emulsion PCR

Steps in green are automated with the use of the Ion Chef
8 samples can be prepared per Ion Chef run

Workflow: Precision ID

Step: PCR amplification

Purpose: Targeted amplification of STRs/SNPs

Target 1 ng of gDNA
23 PCR cycles (STRs)
21 PCR cycles (SNPs)
20 µL volume

Workflow: Precision ID

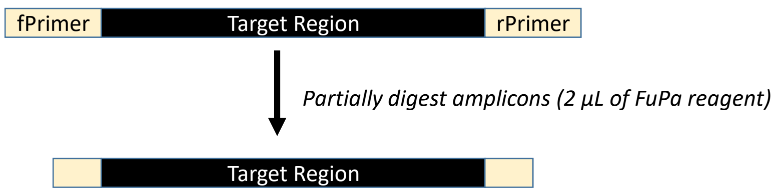
Step: Partially digest PCR products

Purpose: Prepare amplicons for the ligation of adapters

Temperature	Time
50°C	10 minutes
55°C	10 minutes
60°C	20 minutes
10°C	Hold (for up to 1 hour)

What is the FuPa reagent?

The FuPa reagent is a proprietary reagent.
My guess: it partially digests the primer regions and prepares the amplicon for ligation of adapters
Patent https://patents.google.com/patent/US8728728B2/en



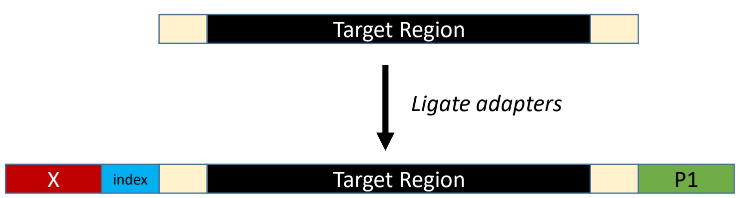
Single base overhang? Phosphorylated 5 ' end ???

Workflow: Precision ID

Step: Ligate adapters to PCR products

Purpose: Attach indices and P1/X sequences

Panel	Temperature	Time
Precision ID GlobalFiler™ NGS STR Panel v2	22°C	30 minutes
	68°C	10 minutes
	10°C	Hold (for up to 24 hours)



Order of addition	Component	Volume
1	Switch Solution (yellow cap)	4 µL
2	Precision ID IonCode™ Barcode Adapter	2 µL
3	DNA Ligase (blue cap)	2 µL
—	**Total volume**	**~30 µL**

You *must* ligate a different barcode adapter to each library

Workflow: Precision ID

Step: Purification of libraries

Purpose: Purify the ligated libraries; remove unincorporated ligation adapters

Add 45 µL (1.5X sample volume) of Agencourt™ AMPure™ XP Reagent to each library.
Mix the bead suspension with the DNA thoroughly, then incubate the mixture for 5 minutes at room temperature.
Place the plate in a magnetic rack, then incubate for 2 minutes
Carefully remove, then discard the supernatant without disturbing the pellet.
Two washes with 70% ethanol
Ensure that all ethanol droplets are removed from the wells.
Keep the plate in the magnet, then air-dry the beads at room temperature for 5 minutes.

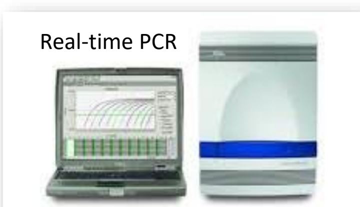
Bead-based purification of the libraries
Removing the initial PCR and ligation reagents, primers

Elute and collect each library…

Workflow: Precision ID

Step: Quantification of libraries (qPCR)

Purpose: Quantify the libraries (prior to pooling)



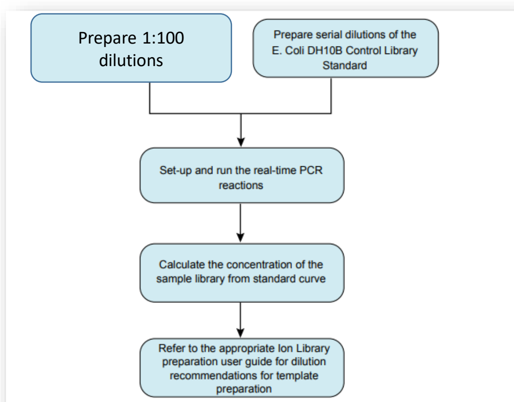


Figure 1 Library quantification with the Ion Library TaqMan® Quantitation Kit

Workflow: Precision ID

Step: Dilute and Pool

Purpose: Prepare equimolar pool for emulsion PCR

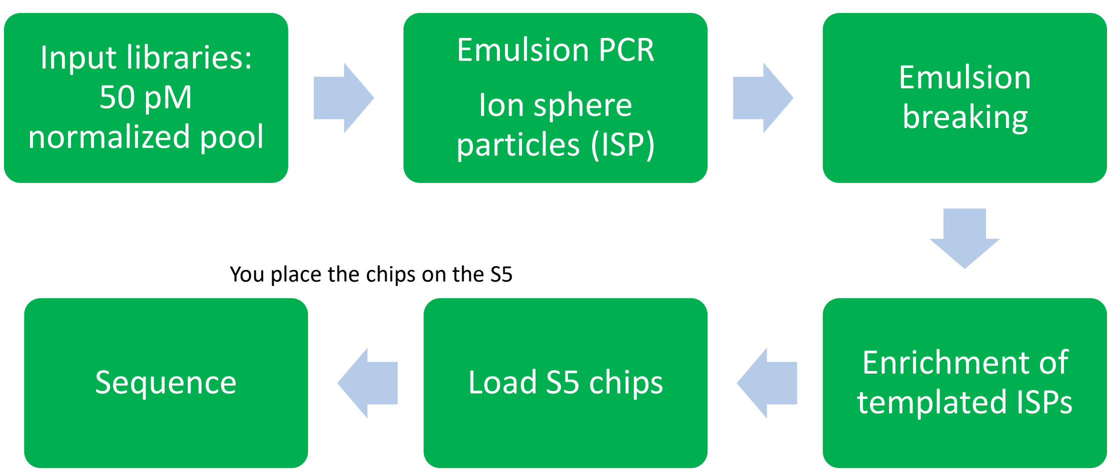
Panel	Dilute to	Minimum volume
Precision ID GlobalFiler™ NGS STR Panel v2	50 pM	25 µL

Panel	Dilute to	Minimum volume
Precision ID Ancestry Panel *or* Precision ID Identity Panel	30 pM	25 µL

If you do not have enough library it is possible to amplify some of the existing library

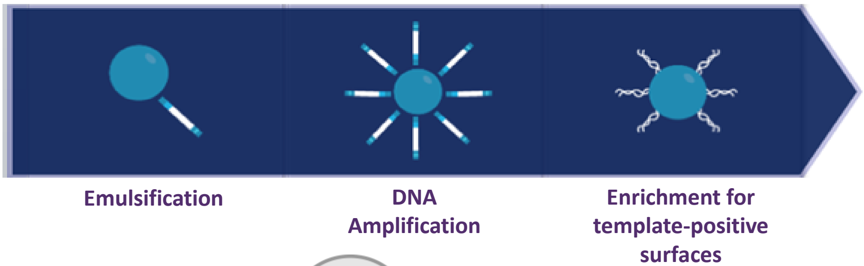
Amp 5-10 cycles
Agencourt™ AMPure™ XP Reagent
First round at 0.5X bead-to-sample-volume ratio: High molecular-weight DNA is bound to beads, while amplicons and primers remain in solution. **Save the supernatant.**
Second round at 1.2X bead-to-original-sample-volume ratio: Amplicons are bound to beads, and primers remain in solution. **Save the bead pellet, and elute the amplicons from the beads.**
Quantify by qPCR

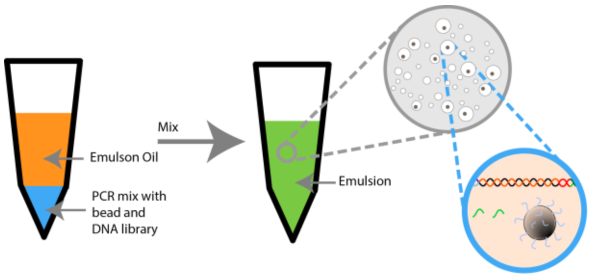
Precision ID panels (Sequenced on S5)



Steps in green are automated with the use of the Ion Chef

Templating Workflow





Slide courtesy of Matt Gabriel – Thermo Fisher

Workflow: Precision ID

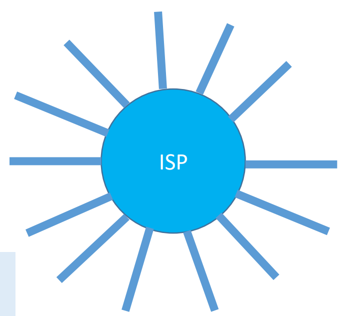
Step: Emulsion PCR

Purpose: Generate clonal "clusters" on a bead (ISP)

ISP = Ion Sphere Particle

The clonal amplification and enrichment steps are performed using the Ion One Touch 2 system or Ion Chef

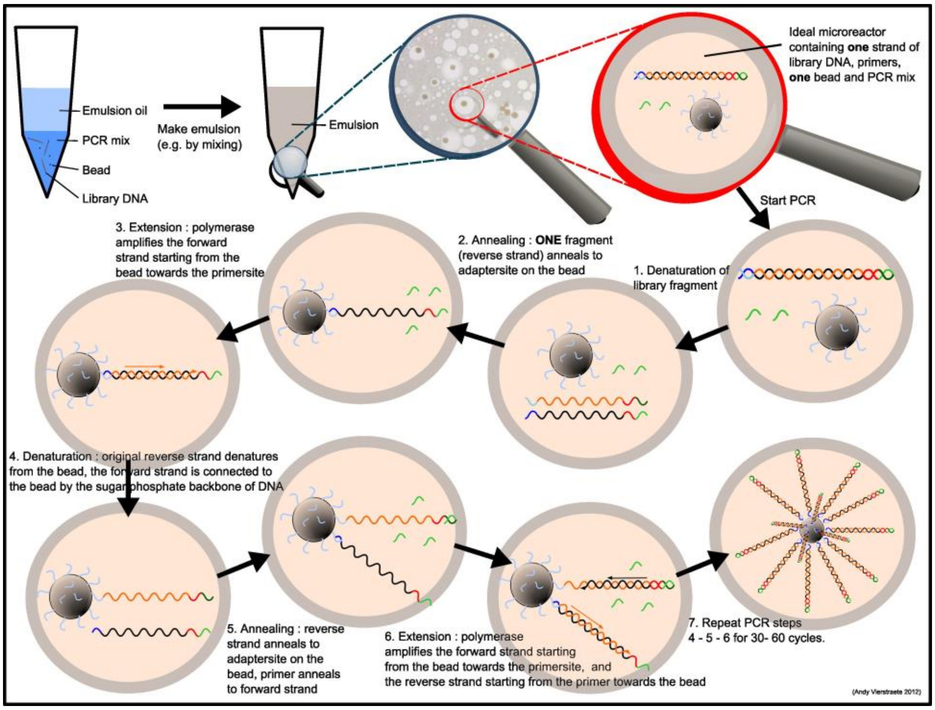
+



ISP covered with Primers that will bind to the P1 adapter

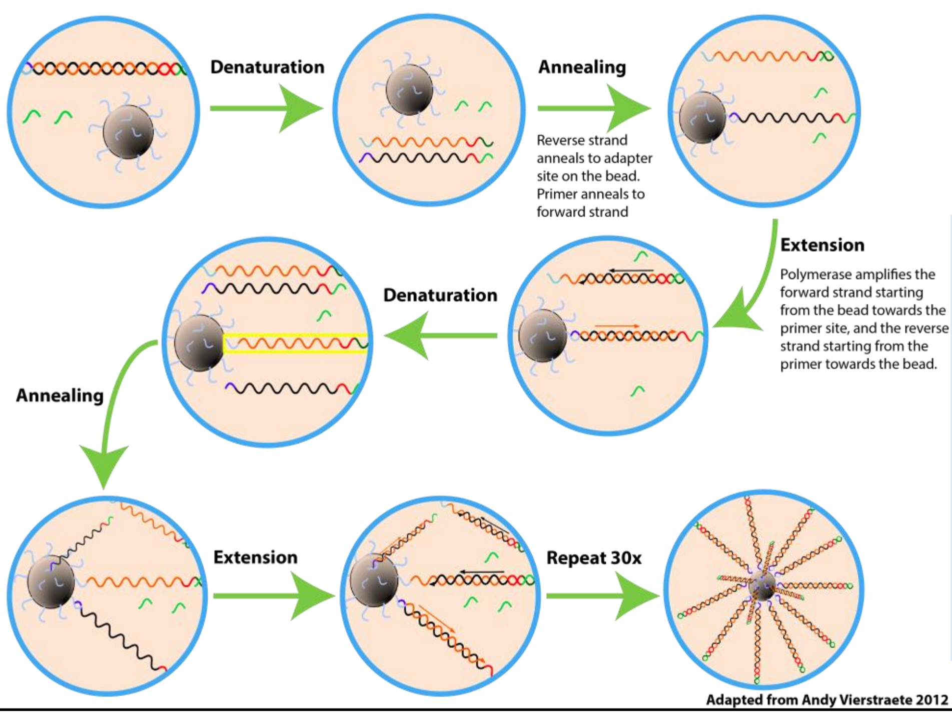
Clonal amplification is performed by an emulsion PCR
Libraries bind to ISPs with from P1 adapter end
Primers (on ISP and in emulsion) amplify from the X and P1 regions

I believe the X primer is biotinylated – for enrichment…later



One strand (forward) is bound to the ISP and the reverse natures/denatures and creates a 'cluster' covering the ISP

Post PCR the emulsions are broken open and the ISPs are 'enriched'



After the DNA strands are amplified, the emulsion from the preceding step is broken using isopropanol and detergent buffer.

The solution is then vortexed, centrifuged, and magnetically separated.

The resulting solution is a suspension of empty, clonal and non-clonal beads, which will be filtered in the next step.

Workflow: Precision ID

Step: Enrichment

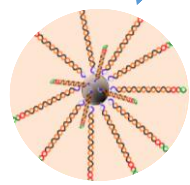
Purpose: Remove ISPs that do not have DNA attached to the surface

The ISPs are no longer in an emulsion droplet

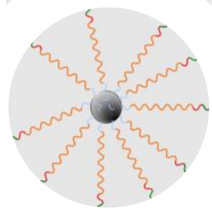
Enrichment
ISPs with clonal libraries will be bound to streptavidin covered magnetic beads (biotinylated X primer)
Clonal ISPs will be captured (an ISP with a single colony of amplified library)

ISPs **without** DNA attached to the surface will be washed away

Poly-clonal ISPs will also be captured
If more than one unique library molecule was present in the emulation PCR reactor
Reagent concentrations are prescribed to minimize this possibility

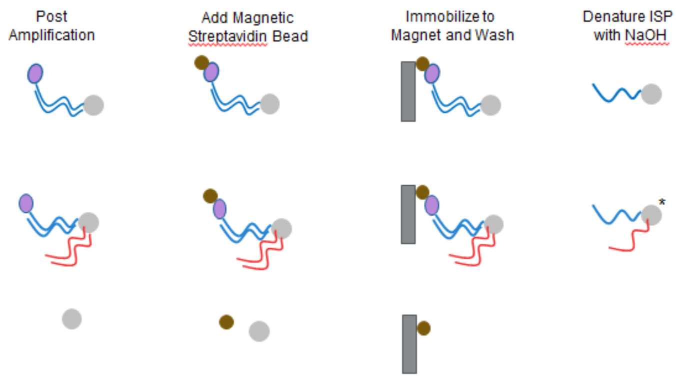


The biotinylated strand will be bound to the magnetic bead
The ISP is covered with ssDNA and ready for sequencing



Enrichment

- Streptavidin beads
- Binds to Adapter X only
  - Template Positive ISPs have Adapter X at the ends



Note: Polyclonality will still occur– this is why library quant is critical

Slide courtesy of Matt Gabriel – Thermo Fisher

Workflow: Precision ID

Step: Load the S5 chip

Purpose: Efficiently load the ISPs onto the chip

Chip Loading
The Ion Chef will load the ISPs onto the chip (reduce air bubbles, even/efficient loading)
You physically load the chips onto the S5 instrument

3-6 M reads per chip
Up to 600 bp

15-20 M reads per chip
Up to 600 bp

60-80 M reads per chip
Up to 200 bp

Workflow: Precision ID

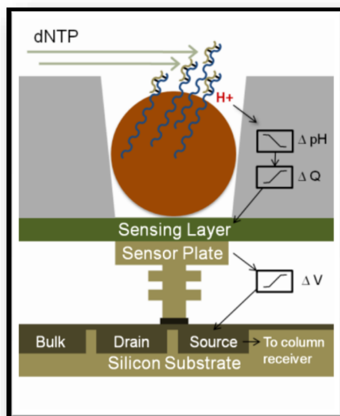
Step: Sequencing

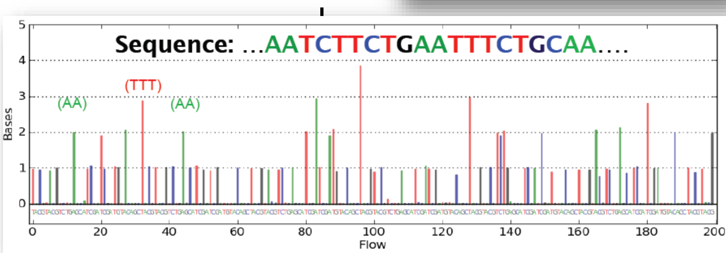
Purpose: Sequence the DNA stranded bound to the ISPs

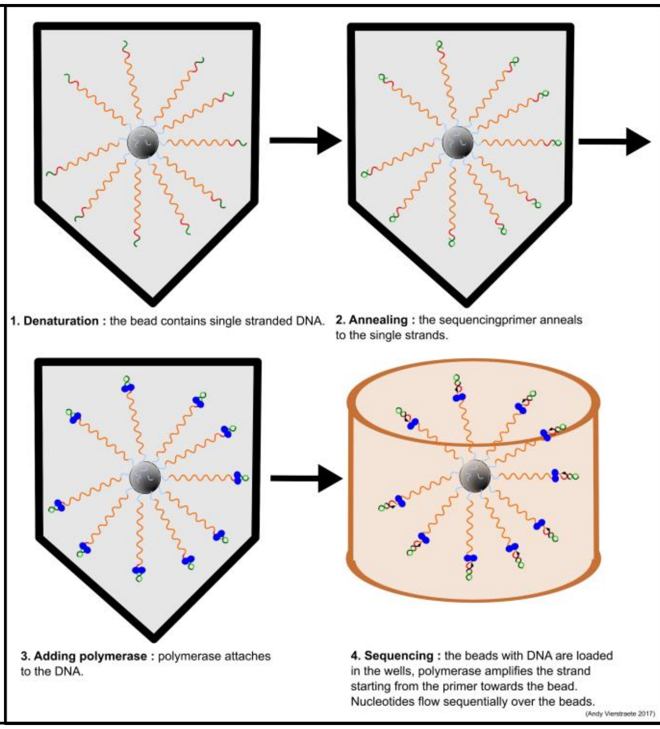
Ion Torrent Sequencing
Semiconductor-based sequencing

Electronically detects polymerase-driven base incorporation without the use of fluorescence

- Chip flooded with **one nucleotide after another**
- H^+ released when a complementary base is added to template
- Charge from the ion causes detectable pH change
- Sequencer calls the base







Ion Torrent Movie

- https://www.youtube.com/watch?v=DyijNS0LWBY

S5 Diagnostics (530 chip Identity SNP panel)

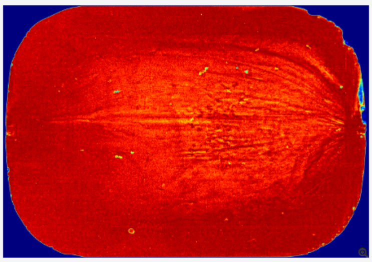
Unaligned Reads

2.02 G
Total Bases

86
Key Signal

90 %
ISP Loading

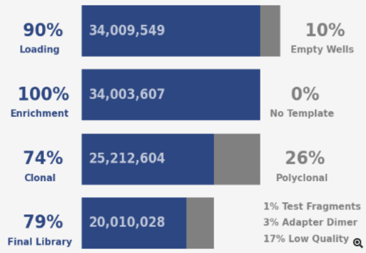
ISP Density



20,010,028
Total Reads

60 %
Usable Sequence

Chip well details

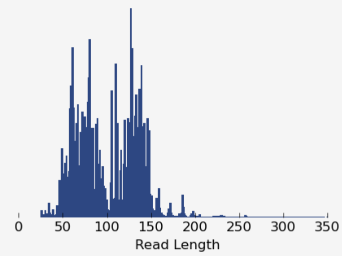


101 bp
Mean

96 bp
Median

126 bp
Mode

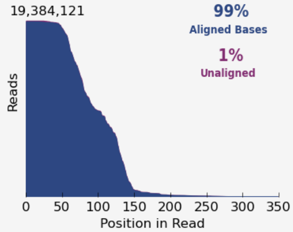
Read Length



Aligned Reads hg19(hg19 from zip)

1.93 G
Total Aligned Bases

0.6X
Reference Coverage

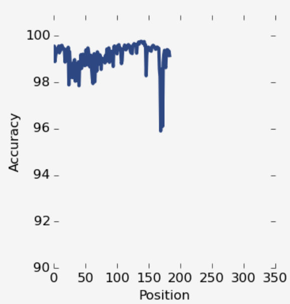


99%
Aligned Bases

1%
Unaligned

	Count	%
Total Reads	19,384,121	—
Aligned Reads	19,310,626	99.6%
Unaligned Reads	73,495	0.4%

99.0%
Mean Raw Accuracy 1x



1.79 G
AQ17 Total Bases

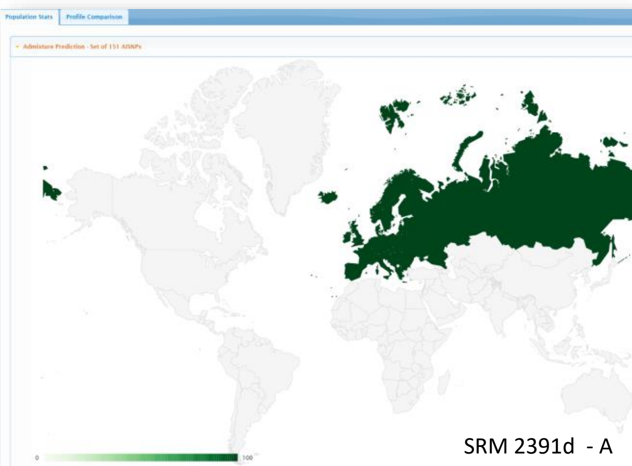
Alignment Quality

	AQ17	AQ20	Perfect
Total Number of Bases [bp]	1.79 G	1.51 G	1.24 G
Mean Length [bp]	97	86	71
Longest Alignment [bp]	337	337	320
Mean Coverage Depth [x]	0.6	0.5	0.4

Data analysis in Converge

- Working with a cloud instance of Converge set up by Thermo Fisher (Thank you!)
- Sequence data files from NIST SRM 2391d (STR, SNP, Mito)
- Intended to give a brief overview of the sequence data

Ancestry SNPs



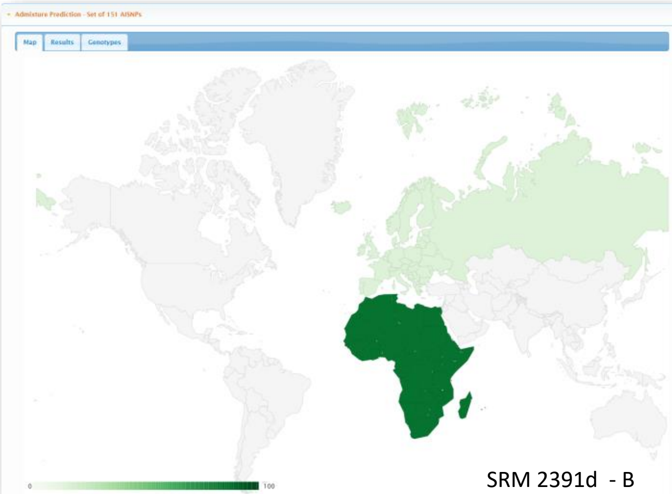
Population Likelihoods - Set of 151 AISNPs

Map | Results | Genotypes

Population	Geo region	Likelihood
Danes	Europe	1.771e-43
Europeans-HapMap	Europe	2.248e-44
Russians	Europe	2.088e-44
Irish	Europe	2.684e-45
Finns	Europe	1.119e-45
Hungarian	Europe	1.088e-45
Komi-Zyrian	Asia	2.023e-46
European Americans	Europe	1.570e-46
Russians, Archangel'sk	Europe	1.394e-46
Adygei	Europe	5.090e-47
Chuvash	Europe	8.939e-48
Jews, Ashkenazi	Europe	5.522e-48
Jews, Sephardic	Europe	1.872e-51
Greeks	Europe	1.792e-51
Druze	Asia	4.390e-54
Mohanna	Asia	3.940e-54
Pashtun	Asia	1.487e-54
Palestinian	Asia	1.047e-55
Sardinian	Europe	7.165e-56

Results from the Torrent Server Suite: HID SNP Genotyper Report

Ancestry SNPs



Population Likelihoods - Set of 151 AISNPs

Map | Results | Genotypes

Population	Geo region	Likelihood
African Americans	Africa	1.395e-44
Chagga	Africa	4.553e-48
Masai	Africa	1.360e-50
Hausa	Africa	1.302e-50
Lisongo	Africa	2.427e-52
Sandawe	Africa	8.246e-53
Zaramo	Africa	8.159e-54
Jews, Ethiopian	Africa	5.201e-55
Ibo	Africa	6.075e-56
Yoruba	Africa	2.118e-56
Yoruba-HapMap	Africa	3.172e-58
Somali	Africa	2.203e-60
Biaka	Africa	1.561e-62
Mbuti	Africa	1.067e-62
Negroid Makrani	Asia	2.701e-63

Results from the Torrent Server Suite: HID SNP Genotyper Report

Precision ID mtDNA Panels

Precision ID mtDNA Whole Genome Panel
Precision ID mtDNA Control Region Panel

Degenerate primers

Degenerate primers account for the high frequency of variants of mtDNA. The number of degenerate primers for each pool is as follows:

Panel	Pool 1	Pool 2	Variant frequencies[1, 2]
Precision ID mtDNA Whole Genome Panel	81 primer pairs 119 degenerates	81 primer pairs 164 degenerates	1000 Genomes: >5% population frequency www.1000genomes.org MitoMap: >700 count www.mitomap.org
Precision ID mtDNA Control Region Panel	7 primer pairs 45 degenerates	7 primer pairs 68 degenerates	

[1] Degenerates were designed to avoid dropouts caused by primer binding SNPs identified from these references.
[2] Additional degenerate primers were added after a round of global customer testing.

Two PCR reactions per sample

Dual panel system and library amplification

A dual panel system covers the mtDNA Genome and Control Region. The following figures are visual representations of the dual panels, Hypervariable Regions, and insert locations of the Precision ID mtDNA Control Region Panel.

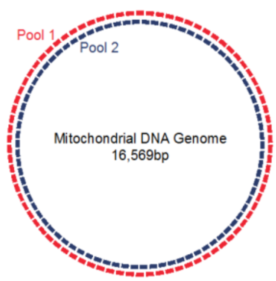


Figure 1 Visual representation of the dual panel system. It does not accurately display the actual number of amplicons.

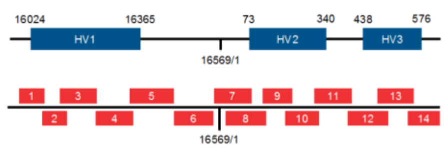
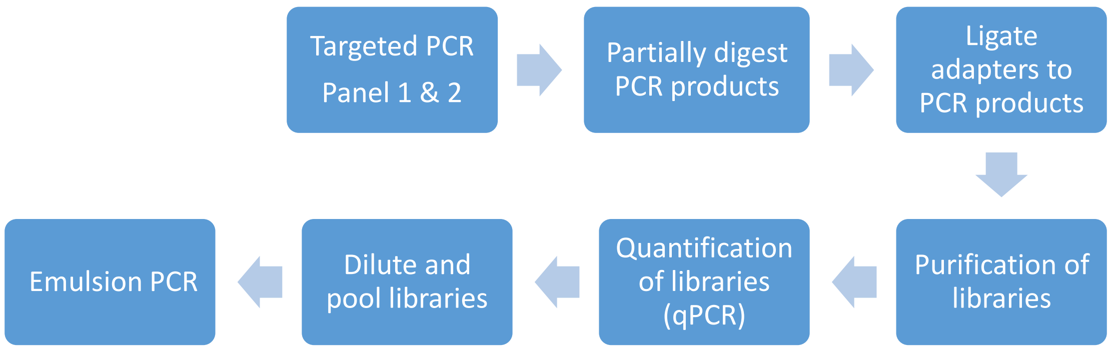


Figure 2 The Hypervariable Regions of the mtDNA Control Region and insert locations of the Precision ID mtDNA Control Region Panel.

Precision ID Panels (Sequenced on S5)



View Mito data in Converge

Summary

- Emulsion PCR is performed to generate clonal library populations on the ISP
  - Not performed on the sequencer
- Sequencing is carried out by flowing successive NTPs onto the chip
- Quantification steps for library normalization
  - qPCR

Forensic Science International: Genetics 29 (2017) 174–180

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

A performance evaluation of Nextera XT and KAPA HyperPlus for rapid Illumina library preparation of long-range mitogenome amplicons

Joseph D. Ring[a,b,*], Kimberly Sturk-Andreaggi[a,b], Michelle A. Peck[a,b], Charla Marshall[a,b]

[a] Armed Forces DNA Identification Laboratory, A Division of the Armed Forces Medical Examiner System, 115 Purple Heart Drive, Dover AFB, DE 19902, United States

[b] ARP Sciences, LLC, Contractor Supporting the Armed Forces Medical Examiner System, 9210 Corporate Boulevard, Suite 120, Rockville, MD 20850, United States

Comparing Nextera XT and KAPA HyperPlus library preparation methods

BMC Genomics

BioMed Central

Methodology article

Open Access

Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences

Liane Fendt[1], Bettina Zimmermann[1], Martin Daniaux[2] and Walther Parson*[1]

Address: [1]Institute of Legal Medicine, Innsbruck Medical University, Müllerstrasse 44, Austria and [2]Clinical Department of Radiology, Innsbruck Medical University, Austria

Email: Liane Fendt - liane.fendt@i-med.ac.at; Bettina Zimmermann - bettina.zimmermann@i-med.ac.at; Martin Daniaux - martin.daniaux@uki.at; Walther Parson* - walther.parson@i-med.ac.at

* Corresponding author

Sanger-based paper
Describes a long range PCR approach (two ≈8.5k amplicons)

Whole Mito Genome (Sequenced on MiSeq)

KAPA Hyper Plus

Targeted PCR → Quantify PCR products → Fragment, end repair, A-tailing → Ligate Illumina adapters to PCR products → Purification of libraries → Quantification of libraries (qPCR) → Dilute and pool libraries

Two PCR Products are amplified

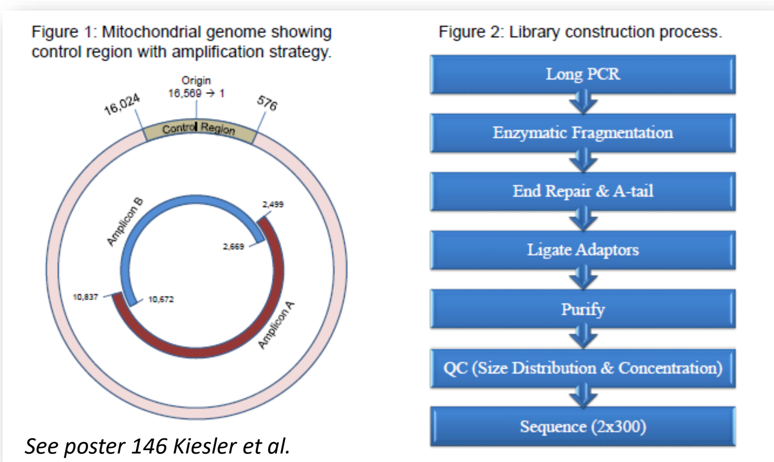


Figure 1: Mitochondrial genome showing control region with amplification strategy.

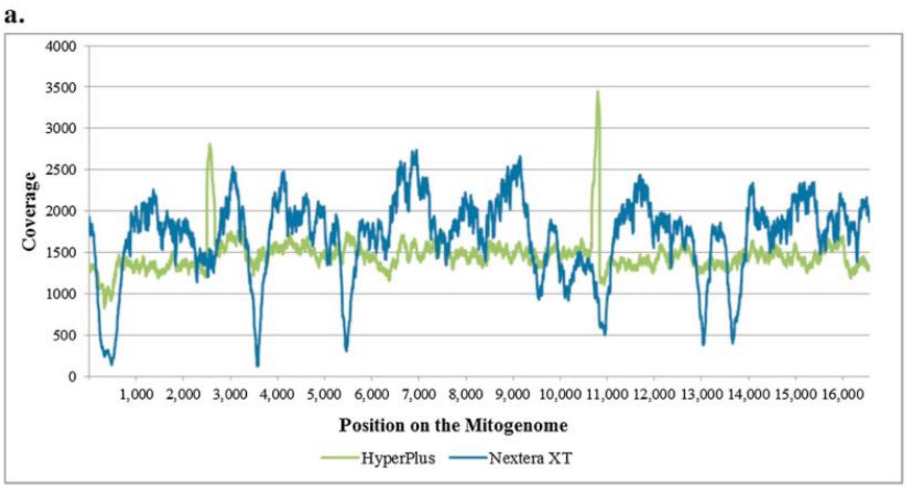
Figure 2: Library construction process.

See poster 146 Kiesler et al.

The large PCR products ≈8.5 kb are fragmented and prepared for Illumina dual index adapters

This approach works well with high quality DNA where large PCR products can be successfully amplified (reference samples)

Coverage balance comparison
KAPA HyperPlus and Illumina Nextera XT



View data in IGV (Whole Mito Genome)

- Mito genomes assembled in GeneMarker HTS (SoftGenetics)

QIAGEN Targeted DNA Panels

QIAseq Targeted DNA Panels

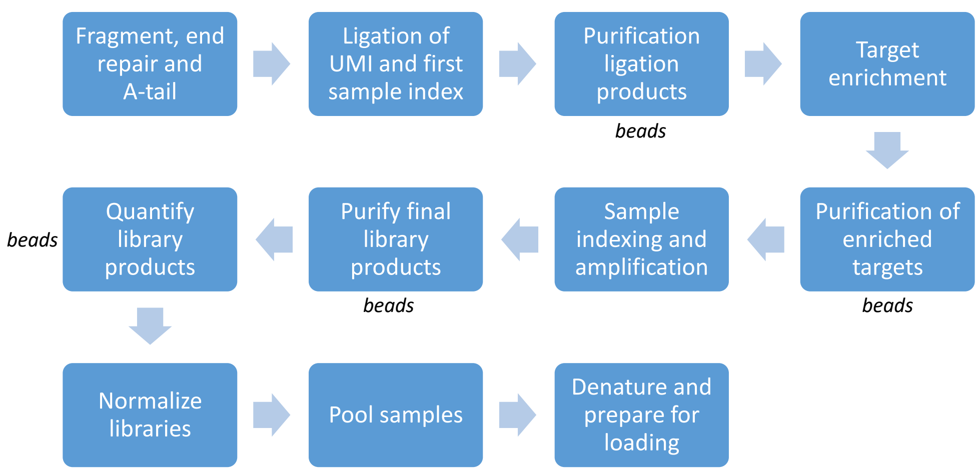
Print

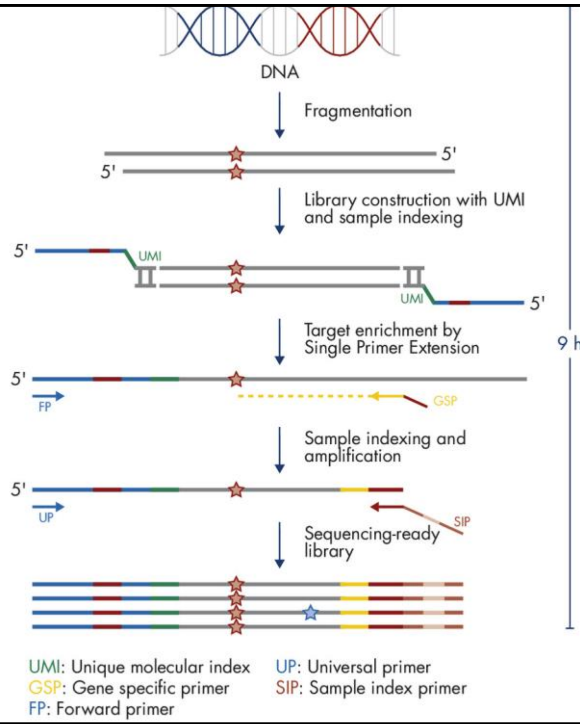
Digital DNA sequencing to confidently detect low-frequency variants

- Digital sequencing enabled by molecular barcodes to remove PCR duplicates
- Complete Sample to Insight solution streamlines the workflow
- Compatibility with low-quality DNA enables efficient sequencing of FFPE and cfDNA samples
- Minimal DNA input to preserve precious samples
- Optimized buffers and conditions to achieve high coverage of GC-rich regions

Product	Product no.	Cat. no.
Human Mitochondria Panel		DHS-105Z

QIAGEN Targeted DNA Panel (Sequenced on MiSeq or S5)



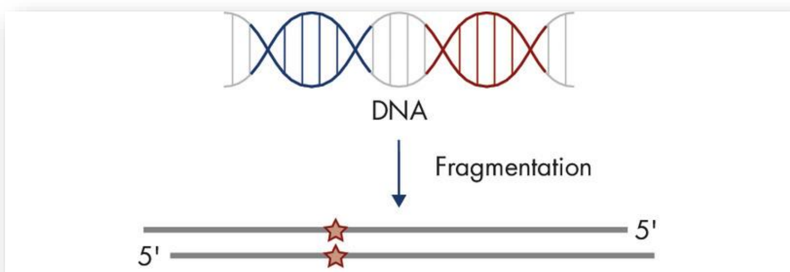


Workflow: QIAGEN Human Mitochondria Panel

Step: Fragmentation, end repair and A-tailing

Purpose: Prepare DNA for ligation

Started with 10 ng of nDNA
≈1600 nuclear diploid copies
Mito copies ≈ 500k
Assume ≈ 300 mtGenomes per cell

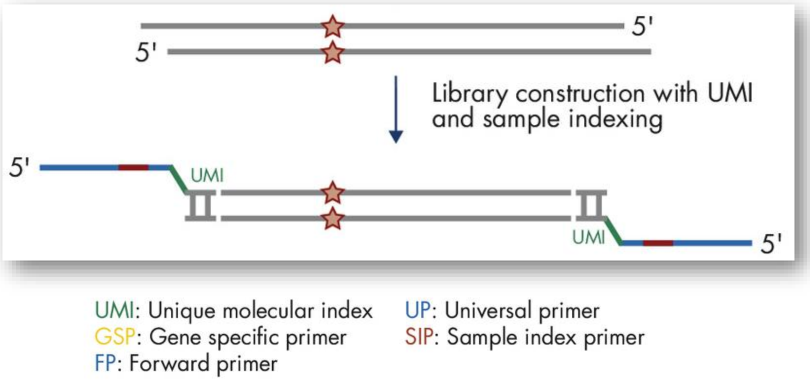


A main point here is that we are starting with a larger amount of DNA template (relative to the other workflows)

Workflow: QIAGEN Human Mitochondria Panel

Step: Adapter ligation

Purpose: Label the pre-amp fragments adapters containing UMIs and the first sample index



Ligation of unique adapters (UMI) and sample index (red) to each piece of fragmented DNA in the solution (no targeted PCR steps yet)

What is a Unique Molecular Index (UMI)?

- Errors in PCR and the sequencing process can lead to higher noise/false positives (is it a low frequency SNP or 'noise'?)

- Enhance low variant detection
- 12-base fully random sequence (4^{12} possible indices per adapter)
- **Must be attached pre-targeted PCR**
- Each fragmented molecule receives a unique UMI sequence

- UMIs are binned informatically and variants can be assessed…

What is a Unique Molecular Index (UMI)?

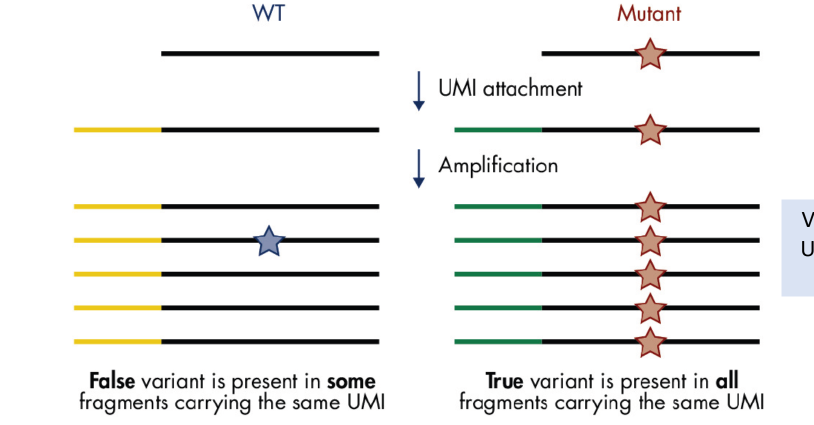


Figure 3. Principle of variant detection with UMIs. Each original molecule is tagged by a UMI. True variants are those mutations present in the majority of reads within a UMI, while false positives are mutations present in only one or a few reads within a UMI. Description of the variant calling algorithm can be found and downloaded from Xu, C, et al. BMC Genomics (2017) **18**, 5.

Variants observed in the same UMI class in the same position can be considered 'true'