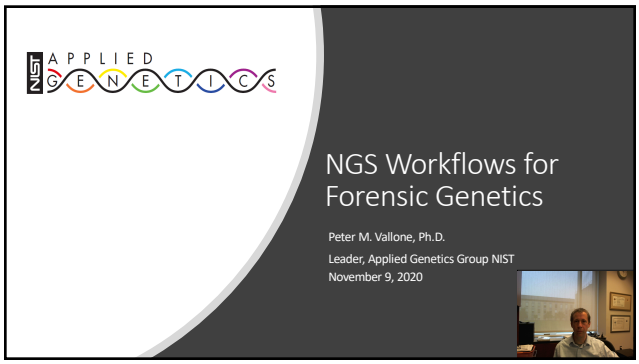
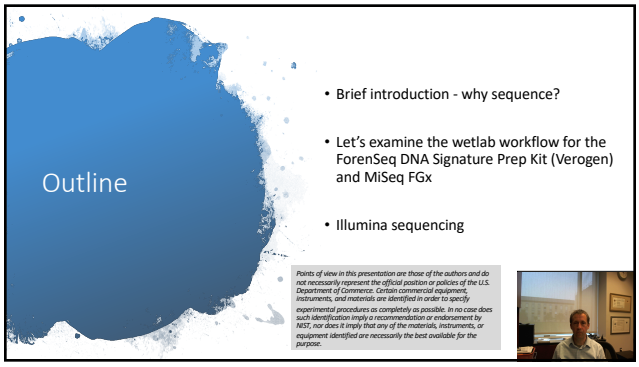
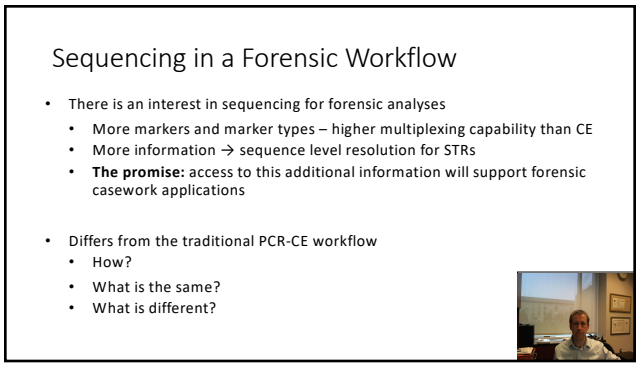

APPLIED GENETICS
NGS Workflows for Forensic Genetics
Peter M. Vallone, Ph.D.
Leader, Applied Genetics Group NIST
November 9, 2020

1


Outline
• Brief introduction - why sequence?
• Let's examine the wetlab workflow for the ForenSeq DNA Signature Prep Kit (Verogen) and MiSeq FGx
• Illumina sequencing
Points of view in this presentation are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

2

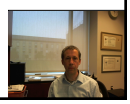

Sequencing in a Forensic Workflow
• There is an interest in sequencing for forensic analyses
• More markers and marker types – higher multiplexing capability than CE
• More information → sequence level resolution for STRs
• **The promise:** access to this additional information will support forensic casework applications
• Differs from the traditional PCR-CE workflow
• How?
• What is the same?
• What is different?

3

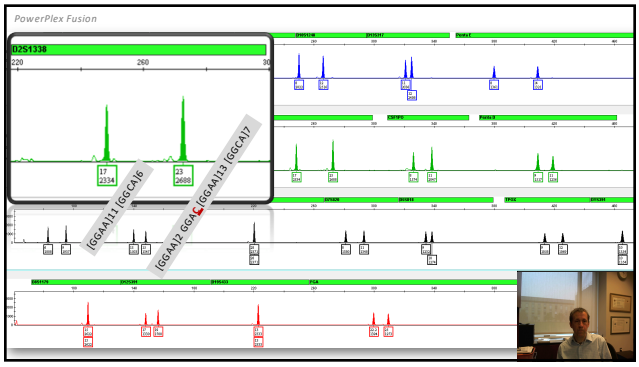
Why sequence STRs?

- Sequencing of STRs
  - STR motif sequence variation; flanking region variation (more polymorphic)
  - Further understand simple versus complex repeat motifs
  - Characterize stutter by sequence
- 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)

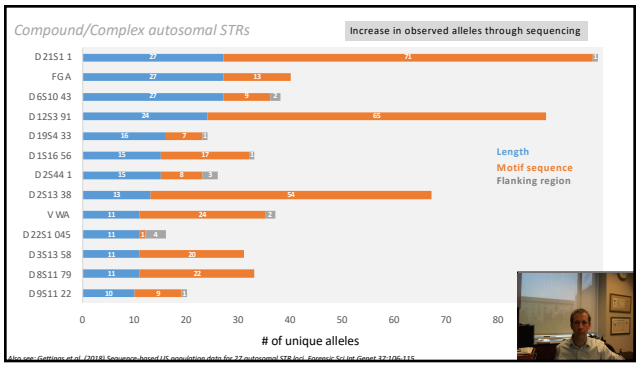
Will still need an interpretation framework
e.g. probabilistic genotyping



4

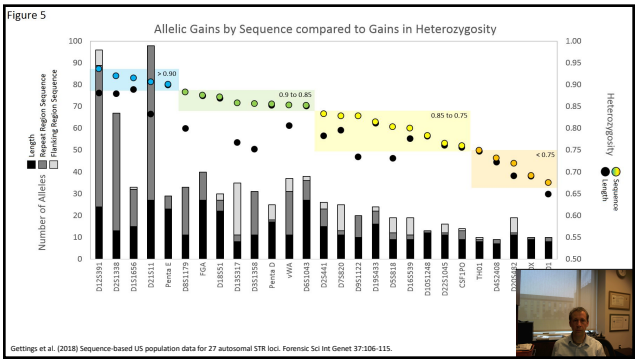


5



6

Figure 5



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

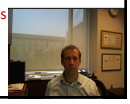
7

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



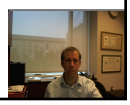
8

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

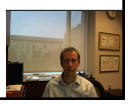
- For now…it seems that forensic genetics will implement a targeted approach



9

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



10

Comparing workflows – targeted sequencing

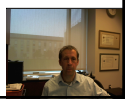
PCR-Capillary Electrophoresis

Collection → Extraction → Quant → PCR → CE → EPG

We'll spend most of our time here today

Collection → Extraction → Quant → PCR → Library Prep → Sequencing → FASTQ

PCR-Sequencing



11

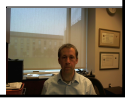
Library preparation

Collection → Extraction → Quant → PCR → Library Prep → Sequencing → FASTQ

Add sequence adapters PCR or ligation → Purify → Quantify → Normalize → Pool

General…we'll get into specifics

Goal: prepare PCR products for sequencing



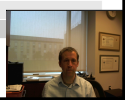
12

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

List not exhaustive – just some common examples

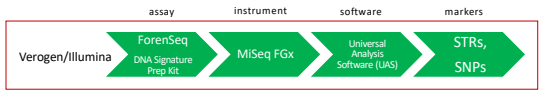
UAS = Universal Analysis Software

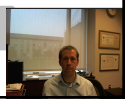


13

For all workflows please visit: https://strbase.nist.gov/pub_pres/ISFG_NGS_Workflows_Workshop_Sept_9.pdf

Example NGS workflows





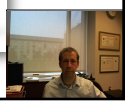
14

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.

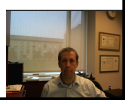


Jäger AC, Alvarez ML, Davis CP, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for targeted next generation sequencing in forensic DNA casework and database laboratories. Forensic Sci Int. 2017;28: 52-70

15

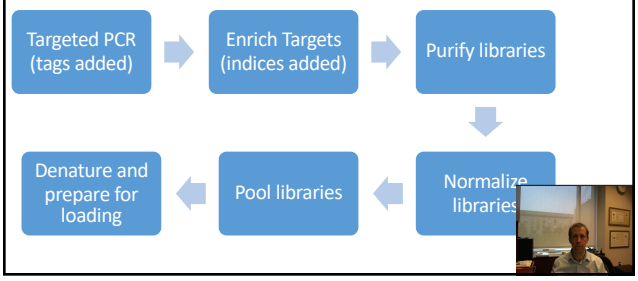
What is the overall goal of library preparation?

- 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

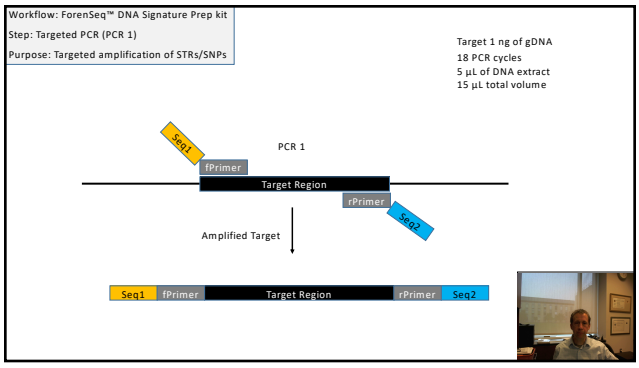


16

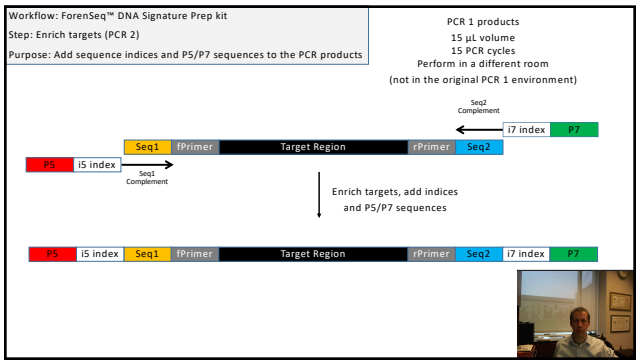
ForenSeq (Sequenced on MiSeq FGx)



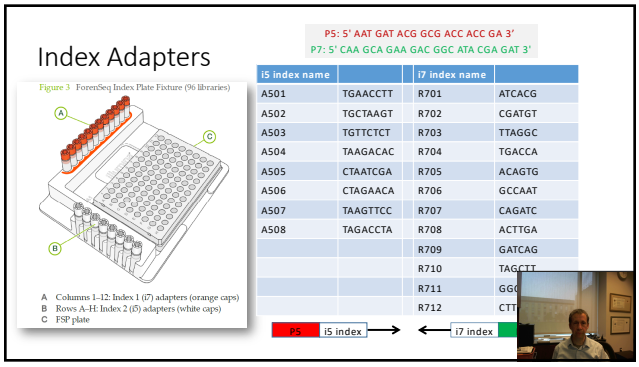
17



18

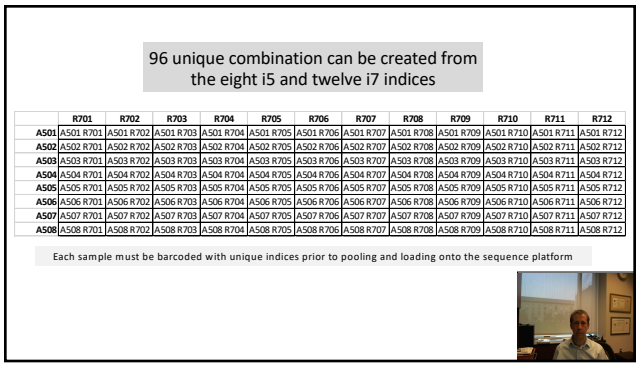


19



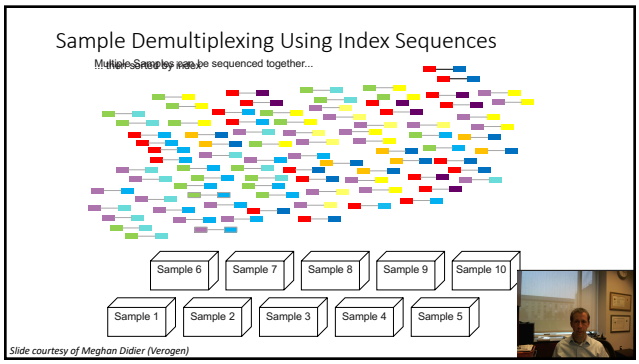
i5 index name		i7 index name	
A501	TGAACCTT	R701	ATCACG
A502	TGCTAAGT	R702	CGATGT
A503	TGTTCTCT	R703	TTAGGC
A504	TAAGACAC	R704	TGACCA
A505	CTAATCGA	R705	ACAGTG
A506	CTAGAACA	R706	GCCAAT
A507	TAAGTTCC	R707	CAGATC
A508	TAGACCTA	R708	ACTTGA
		R709	GATCAG
		R710	TAGCTT
		R711	GGC
		R712	CTT

20

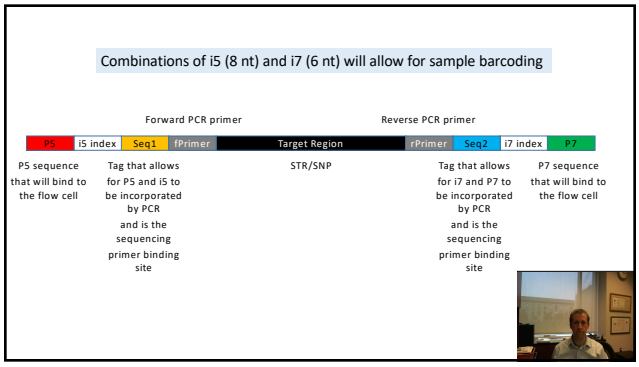


	R701	R702	R703	R704	R705	R706	R707	R708	R709	R710	R711	R712
A501	A501 R701	A501 R702	A501 R703	A501 R704	A501 R705	A501 R706	A501 R707	A501 R708	A501 R709	A501 R710	A501 R711	A501 R712
A502	A502 R701	A502 R702	A502 R703	A502 R704	A502 R705	A502 R706	A502 R707	A502 R708	A502 R709	A502 R710	A502 R711	A502 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

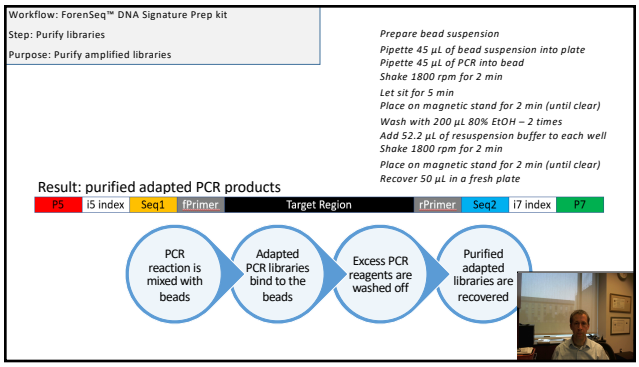
21


Sample Demultiplexing Using Index Sequences
Multiple Samples can be sequenced together...
...then sorted by index
Sample 6
Sample 7
Sample 8
Sample 9
Sample 10
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Slide courtesy of Meghan Didier (Verogen)

22

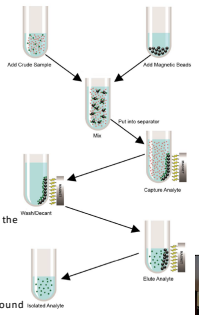

Combinations of i5 (8 nt) and i7 (6 nt) will allow for sample barcoding
Forward PCR primer
Reverse PCR primer
P5
i5 index
Seq1
fPrimer
Target Region
rPrimer
Seq2
i7 index
P7
P5 sequence that will bind to the flow cell
Tag that allows for P5 and i5 to be incorporated by PCR and is the sequencing primer 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

23


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
P5
i5 index
Seq1
fPrimer
Target Region
rPrimer
Seq2
i7 index
P7
PCR reaction is mixed with beads
Adapted PCR libraries bind to the beads
Excess PCR reagents are washed off
Purified adapted libraries are recovered

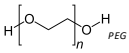
24

Beads! Beads! Beads!



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

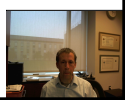
Polystyrene core surrounded by a layer of magnetite coated with carboxyl molecules



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



25

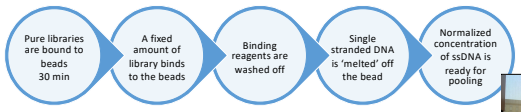
Workflow: ForenSeq™ DNA Signature Prep kit

Step: Normalize libraries

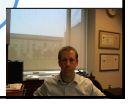
Purpose: Ensure that libraries of varying yields are equally represented within the sequencing run

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



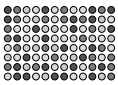
Approx. 2 nM of library post bead normalization



26

ForenSeq Library Preparation

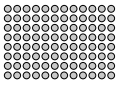
Bead Normalization



Purified libraries:
Range of yields

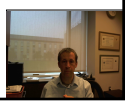
Bead-based Normalization

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



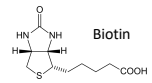
Slide courtesy of Meghan Didier (Verogen)

27

Streptavidin and Biotin

Streptavidin (SA)

+



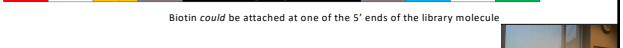
Biotin

Dissociation constant (Kd) ≈ 10^{-14}

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

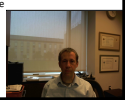
The SA-Biotin complex is stable to organic solvents, denaturants, detergents, temperature, and pH

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



28

Workflow: ForenSeq™ DNA Signature Prep kit

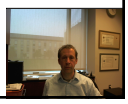
Step: Pool libraries

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 an 8 strip tube – then pooled into a single tube

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

flow cell

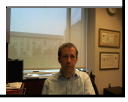


29

MiSeq FGx Sequencing

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

Slide courtesy of Meghan Didier (Verogen)



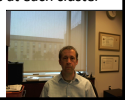
30

Workflow: ForenSeq™ DNA Signature Prep kit
Step: Sequencing
Purpose: Sequence the PCR products

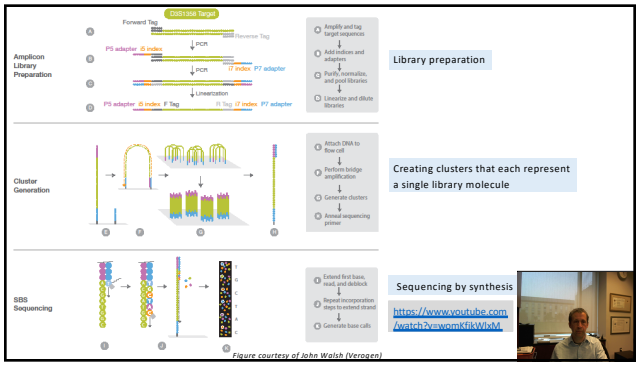
Note: we will not be covering how to set up a specific instrument, loading the system, operational software, etc. This is covered in training materials, software 'wizards'

(Fluorescent) Sequencing by synthesis

- The library you created is hybridized to a flow cell
- Individual strands create 'clusters' through bridge amplification
- Sequencing proceeds one base per cycle
- 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



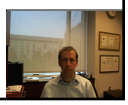
31



32

Summary

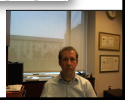
- ForenSeq uses PCR tagging to incorporate library adapters
- Bead-based normalization is used to ensure that each library is of a similar concertation prior to sequencing
- Care should be taken with PCR 2 and i5/i7 indices to avoid contamination
- Sequencing by synthesis – all four bases are incorporated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



33

Acknowledgments

- Verogen
  - Meghan Didier
  - Melissa Kotkin
- NIST
  - Kevin Kiesler
  - Sarah Riman
  - Becky Steffen
  - Katherine Gettings
  - Lisa Borsuk



34