Consultation on Global Trends of HIV Drug Resistance

Session 3: New Technologies for HIVDR Testing

Next Generation Sequencing for HIVDR

Christos J Petropoulos, PhD
Monogram Biosciences
Laboratory Corporation of America® Holdings
South San Francisco, CA, USA

Monogram BIOSCIENCES
LabCorp Specialty Testing Group

What is Next Generation Sequencing?

A high-throughput method used to determine a portion of the nucleotide sequence of an individual's [viral] genome.

This technique utilizes DNA sequencing technologies that are capable of processing multiple DNA sequences in parallel.

Also called massively parallel sequencing and NGS.

http://www.cancer.gov/publications/dictionaries/genetics-dictionary?cdrid=763024

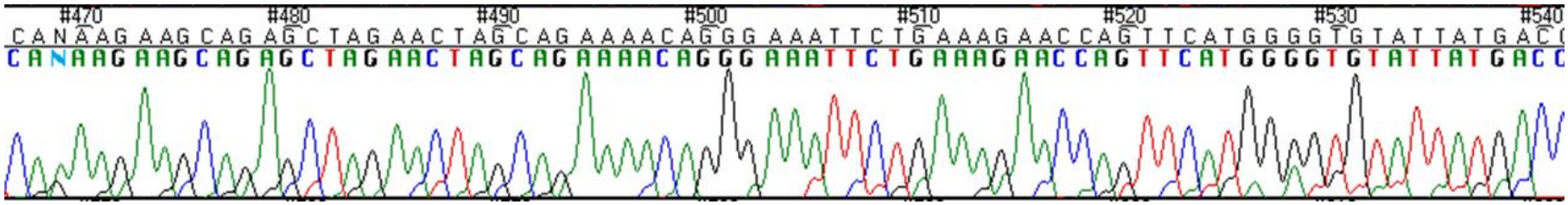
Can be thought of as digital sequencing vs analog sequencing

Why NGS vs Conventional Sequencing?

- NGS data are “clonal” vs Sanger “population” sequences
  - No mixtures in individual sequence reads
  - Individual variants can be quantitated (K65R = 16% vs. K65K/R)
  - Avoids ambiguous translation artifacts
    - e.g. Sanger sequencing of a mixture of S132 (A-G-T) plus H132 (C-A-T) produces an M-R-T codon, which translates as S (A-G-T), N (A-A-T), H (C-A-T) and R (C-G-T); and is reported as S132S/H/N/R
- NGS utilizes same PCR products as conventional Sanger assays
- NGS provides universal sequencing methodology
  - *“If it can be amplified, it can be sequenced”*
  - Eliminates virus/target/subtype specific primers or sequencing
- Accurate (objective/automated), rapid, universal, variant calling

Why NGS Now?

- Originally planned to validate GenoSure Archive using conventional sequencing
- During development it became clear that GenoSure Archive samples often could not be analyzed due to poor sequence quality:



- The cause was determined to be the presence of mixtures of APOBEC-induced hypermutated and non-hypermutated HIV sequences
  - Hypermutated (HM) sequences contain a unusually high percentage of adenine bases which results in differential electrophoretic mobility during capillary electrophoreses leading to poor sequence quality
  - HM results in an overabundance of mutations resulting in stop codons
    - TGG → TAG or TGA or TAA)
- Next generation (clonal) sequencing overcame this technical limitation

Next Generation Sequencing Platforms

- 454 Sequencing / Roche
  - GS Junior System
  - GS FLX+ System
- Illumina (Solexa)
  - HiSeq System
  - Genome analyzer IIx
  - MySeq
- Applied Biosystems - Life Technologies
  - SOLiD 5500 System
  - SOLiD 5500xl System
- Ion Torrent - Life Technologies
  - Personal Genome Machine (PGM)
  - Proton
- Helicos
  - Helicos Genetic Analysis System
- Pacific Biosciences
  - PacBio RS
- Oxford Nanopore Technologies
  - GridION System
  - MinION

Next Generation Sequencing
Amplified Single Molecule Sequencing

Third Generation Sequencing,
Next Next Generation Sequencing,
Single Molecule Sequencing

http://users.ugent.be/~avierstr/nextgen/Next_generation_sequencing_web.pdf

NGS Platforms: Advantages and Limitations

Platform	Why?	Why not?
Roche 454 GS FLX, Junior (emulsion PCR)	Longer read lengths Established platform	Sunsetting platform Homopolymers
Ion Torrent PGM, Proton (emulsion PCR)	Short run times Long read lengths	Homopolymers
Illumina MiSeq (sequence by synthesis)	Superior data quality	Long run times Short read lengths
PacBio RS, Sequel (single molecule sequencing)	Longest read lengths	High start-up and maintenance costs

NGS at Monogram

- Illumina Nextera XT sample preparation
  - 1 ng input DNA requirement
  - Transposome-mediated fragmentation (Tagmentation)
  - Barcode and multiplex up to 96 samples/run
- Illumina MiSeq platform
  - 2x150 bp paired end reads (2x250 bp also available)
  - Run time ~ 24 hours
  - Run size = 96 (94 samples + 2 controls)
  - 24-30 million paired end reads/run (4-6 gB of data)
  - High quality data (>90% of bases higher than Q30)
  - Coverage averages 10,000X (minimum acceptance 1000X)
  - Variant detection down to 1% (given sufficient coverage)
- Custom, universal data analysis pipeline

Sample Preparation: Nextera XT Workflow

Quantitate DNA

↓

Dilute DNA to ~0.2 ng/µL

↓

Tagmentation

↓

Limited cycle PCR to add barcodes and sequencing primer sequences

↓

PCR clean up (Ampure XP)

↓

Library Normalization

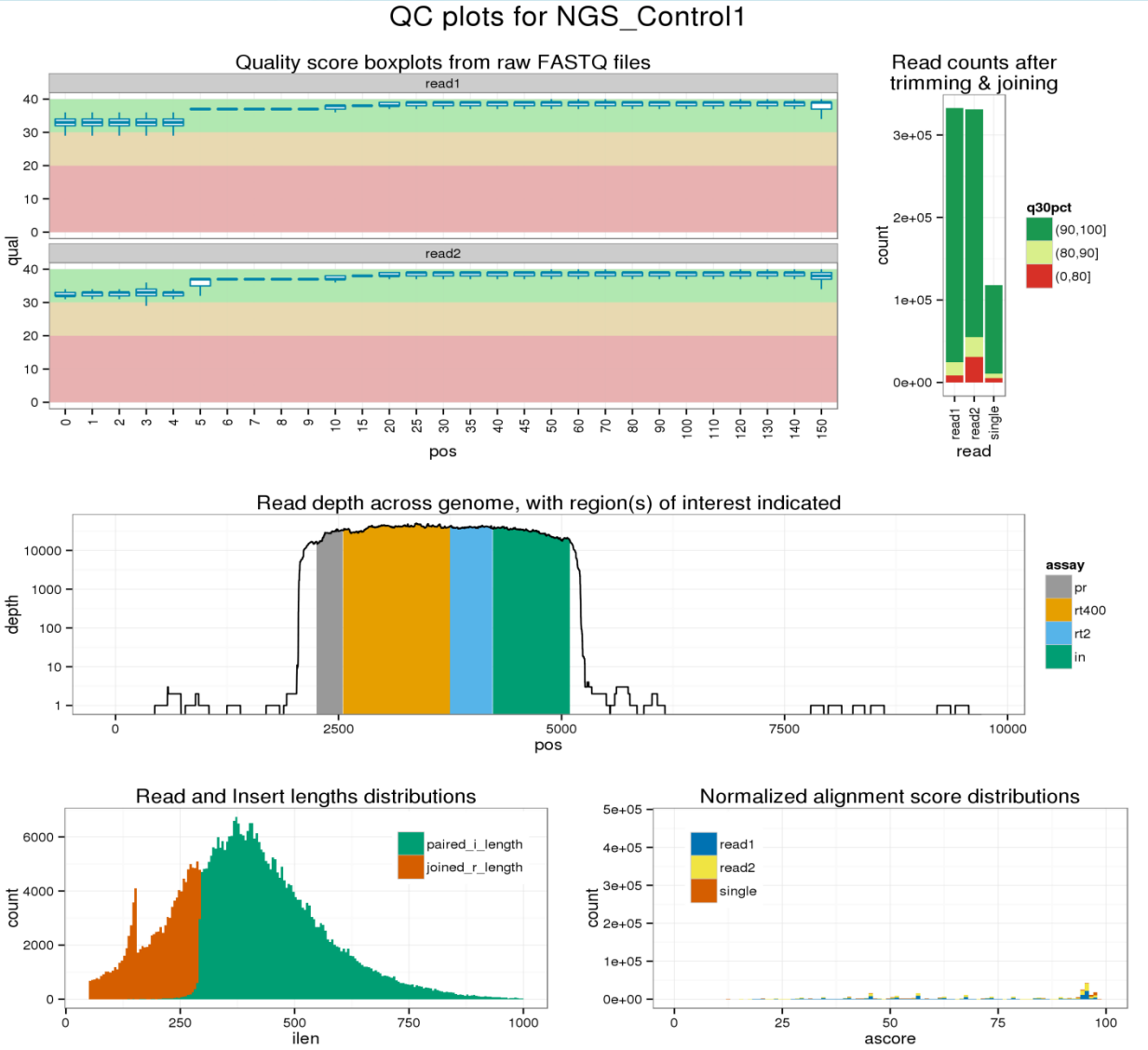
↓

Library Pooling

NGS Data Analysis: Custom Universal Pipeline

- **Reference Selection:** Determines virus species, subtype, and gene region; assigns “best reference”.
- **Alignment:** Performs quality trimming, paired-end joining, and codon-aware alignment to reference.
- **Inspection:** Generates read and alignment quality control statistics.
- **Variant Analysis:** Reports SNP and amino acid variants, tabulates results in a codon-by-codon manner.
- **Summary:** Removes low frequency and/or low quality variants and outputs concise results.
    - Applies G to A hyper-mutation filter to HIV proviral DNA samples

NGS File Output: Inspection Plot



NGS File Outputs: “Mutation_table_aa”

15_117344_1GN.0_10pct_mutation_table_aa.txt [Read-Only]

	A	B	C	D	E	F	G	H	I	J	K
1	assay	cds	pos	aapos	refaa	varaa	mut	aafreq	sample	ref	
2	pr	pr	2280	10	L	I	L10I	0.6621	15_117344_1GN.0	HIV.NL43I151V.KM390026	
3	pr	pr	2289	13	I	V	I13V	0.9935	15_117344_1GN.0	HIV.NL43I151V.KM390026	
4	pr	pr	2349	33	L	V	L33V	0.9934	15_117344_1GN.0	HIV.NL43I151V.KM390026	
5	pr	pr	2361	37	N	S	N37S	0.4431	15_117344_1GN.0	HIV.NL43I151V.KM390026	
6	pr	pr	2436	62	I	V	I62V	0.9948	15_117344_1GN.0	HIV.NL43I151V.KM390026	
7	pr	pr	2439	63	L	Q	L63Q	0.4647	15_117344_1GN.0	HIV.NL43I151V.KM390026	
8	pr	pr	2442	64	I	V	I64V	0.9973	15_117344_1GN.0	HIV.NL43I151V.KM390026	
9	rt400	rt	2652	35	V	I	V35I	0.9074	15_117344_1GN.0	HIV.NL43I151V.KM390026	
10	rt400	rt	2853	102	Q	K	Q102K	0.9873	15_117344_1GN.0	HIV.NL43I151V.KM390026	
11	rt400	rt	3033	162	C	S	C162S	0.9926	15_117344_1GN.0	HIV.NL43I151V.KM390026	
12	rt400	rt	3060	171	F	Y	F171Y	0.9953	15_117344_1GN.0	HIV.NL43I151V.KM390026	
13	rt400	rt	3147	200	T	N	T200N	0.9912	15_117344_1GN.0	HIV.NL43I151V.KM390026	
14	rt400	rt	3180	211	R	K	R211K	0.9916	15_117344_1GN.0	HIV.NL43I151V.KM390026	
15	rt400	rt	3261	238	K	R	K238R	0.1532	15_117344_1GN.0	HIV.NL43I151V.KM390026	
16	rt400	rt	3378	277	R	K	R277K	0.9921	15_117344_1GN.0	HIV.NL43I151V.KM390026	
17	rt400	rt	3396	283	L	I	L283I	0.9212	15_117344_1GN.0	HIV.NL43I151V.KM390026	
18	rt400	rt	3519	324	D	E	D324E	0.9948	15_117344_1GN.0	HIV.NL43I151V.KM390026	
19	rt400	rt	3549	334	Q	C	Q334C	0.9911	15_117344_1GN.0	HIV.NL43I151V.KM390026	
20	rt400	rt	3561	338	T	S	T338S	0.9936	15_117344_1GN.0	HIV.NL43I151V.KM390026	
21	rt400	rt	3621	358	K	R	K358R	0.8549	15_117344_1GN.0	HIV.NL43I151V.KM390026	

15_117344_1GN.0_10pct_mutation_

NGS File Outputs: “Mutation_table_codon”

15_117344_1GN.0_10pct_mutation_table_codons.txt [Read-Only]

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	assay	cds	pos	aapos	refcod	varcod	refaa	varaa	mut	codfreq	aafreq	sample	ref
2	pr	pr	2256	2	CAG	CAA	Q	Q	Q2Q	0.7851	0.9962	15_117344_1GN.0	HIV.NL43I151V.KM390026
3	pr	pr	2271	7	CAG	CAA	Q	Q	Q7Q	0.9947	0.9962	15_117344_1GN.0	HIV.NL43I151V.KM390026
4	pr	pr	2280	10	CTC	ATC	L	I	L10I	0.6613	0.6621	15_117344_1GN.0	HIV.NL43I151V.KM390026
5	pr	pr	2289	13	ATA	GTA	I	V	I13V	0.9918	0.9935	15_117344_1GN.0	HIV.NL43I151V.KM390026
6	pr	pr	2307	19	TTA	CTA	L	L	L19L	0.9871	0.9948	15_117344_1GN.0	HIV.NL43I151V.KM390026
7	pr	pr	2310	20	AAG	AAA	K	K	K20K	0.0319	0.9929	15_117344_1GN.0	HIV.NL43I151V.KM390026
8	pr	pr	2349	33	TTA	GTA	L	V	L33V	0.991	0.9934	15_117344_1GN.0	HIV.NL43I151V.KM390026
9	pr	pr	2352	34	GAA	GAG	E	E	E34E	0.0357	0.9925	15_117344_1GN.0	HIV.NL43I151V.KM390026
10	pr	pr	2361	37	AAT	AGT	N	S	N37S	0.4414	0.4431	15_117344_1GN.0	HIV.NL43I151V.KM390026
11	pr	pr	2394	48	GGG	GGA	G	G	G48G	0.0329	0.9758	15_117344_1GN.0	HIV.NL43I151V.KM390026
12	pr	pr	2406	52	GGT	GGC	G	G	G52G	0.0836	0.9913	15_117344_1GN.0	HIV.NL43I151V.KM390026
13	pr	pr	2424	58	CAG	CAA	Q	Q	Q58Q	0.9876	0.9944	15_117344_1GN.0	HIV.NL43I151V.KM390026
14	pr	pr	2436	62	ATA	GTA	I	V	I62V	0.9926	0.9948	15_117344_1GN.0	HIV.NL43I151V.KM390026
15	pr	pr	2439	63	CTC	TTA	L	L	L63L	0.0252	0.5308	15_117344_1GN.0	HIV.NL43I151V.KM390026
16	pr	pr	2439	63	CTC	CTA	L	L	L63L	0.4998	0.5308	15_117344_1GN.0	HIV.NL43I151V.KM390026
17	pr	pr	2439	63	CTC	CAA	L	Q	L63Q	0.4631	0.4647	15_117344_1GN.0	HIV.NL43I151V.KM390026
18	pr	pr	2442	64	ATA	GTA	I	V	I64V	0.9958	0.9973	15_117344_1GN.0	HIV.NL43I151V.KM390026
19	pr	pr	2448	66	ATC	ATT	I	I	I66I	0.0315	0.995	15_117344_1GN.0	HIV.NL43I151V.KM390026
20	pr	pr	2451	67	TGC	TGT	C	C	C67C	0.988	0.9935	15_117344_1GN.0	HIV.NL43I151V.KM390026
21	pr	pr	2469	73	GGT	GGA	G	G	G73G	0.0396	0.9505	15_117344_1GN.0	HIV.NL43I151V.KM390026
22	pr	pr	2469	73	GGT	GGG	G	G	G73G	0.0285	0.9505	15_117344_1GN.0	HIV.NL43I151V.KM390026
23	pr	pr	2478	76	TTA	CTA	L	L	L76L	0.0365	0.9965	15_117344_1GN.0	HIV.NL43I151V.KM390026
24	pr	pr	2499	83	AAC	AAT	N	N	N83N	0.0336	0.9885	15_117344_1GN.0	HIV.NL43I151V.KM390026

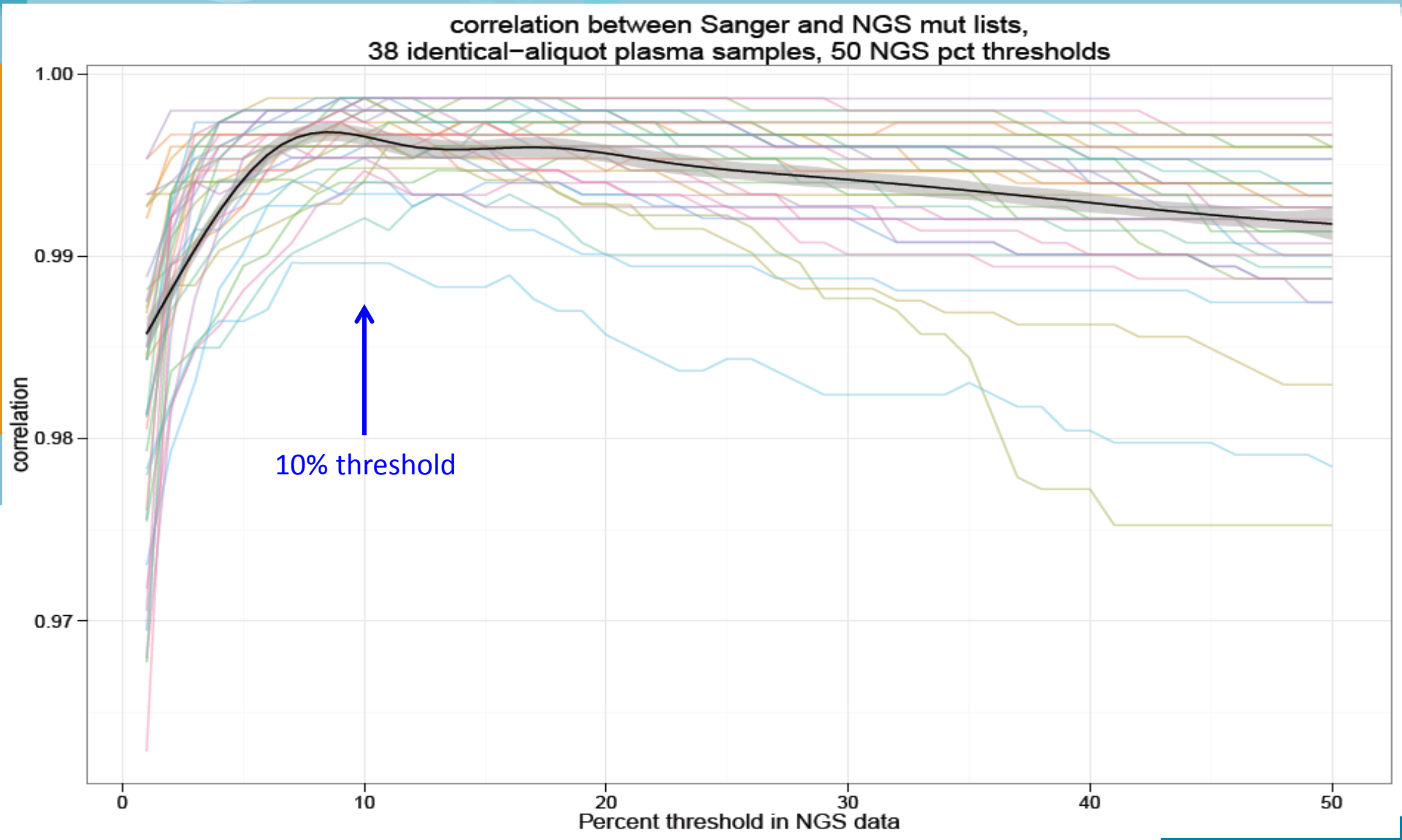
15_117344_1GN.0_10pct_mutation_

"Quantitative Sequencing"

accession	corelation coefficient
14-118298	99.53%
14-118534	99.87%
14-118637	99.00%
14-118716	99.67%
14-118321	99.87%
14-116944	99.47%
14-118624	99.33%
14-119249	99.73%
14-118515	99.14%
14-118519	99.73%
14-118694	99.14%
14-118553	99.87%
14-118670	99.21%
14-117545	99.60%
14-120149	99.73%
14-118003	99.60%
14-118713	99.87%
14-118247	99.67%
14-118641	99.73%
14-118658	99.87%
14-118693	99.27%
14-119723	99.47%
14-119730	99.73%
14-119743	99.53%
14-119745	99.87%
14-118737	99.87%
14-119726	99.47%
14-119855	99.67%
14-118461	99.54%
14-118339	99.47%
14-118325	99.07%
14-120075	99.80%
14-118827	99.73%
14-118810	99.87%
14-118456	99.47%
14-118573	99.53%
14-118181	99.87%

- Replicates the sensitivity to detect variants using Sanger sequencing-based drug resistance assays
  - Applies a 10% variant detection threshold
  - Not "deep sequencing" or minor variant detection
- ≥ 99% correlation between Sanger (GS-PRIme) and NGS mutation lists at a 10% variant threshold
  - Comparison included 38 clinical samples
  - Assessment included 787 AA positions across the HIV-1 PR/RT/IN coding regions
- Improves sequencing accuracy within regions of high variability and/or length polymorphisms
  - e.g. HIV *gag/env*, HCV NS5A
- *Used for routine HCV DR and HIV GS-Archive testing*

Bridging NGS and Sanger Data:



HIV DNA (blood) vs HIV RNA (plasma)?

Drug Resistance Profiles Derived from HIV-1 DNA in ARV Suppressed Patients Correlate with Historical Resistance Profiles Obtained from HIV-1 Plasma RNA

J. Toma[1], Y. Tan[1], S. Cai[1], O. Solberg[1], W. Huang[1], C. Walworth[1], J. M. Whitcomb[1], J. Martin[2], S. G. Deeks[2], C. J. Petropoulos[1]

[1]Monogram Biosciences, South San Francisco, CA

[2]Department of Medicine, San Francisco General Hospital, University of California San Francisco, San Francisco, CA

ICAAC/ICC 2015, September 17-21, San Diego, CA

Monogram BIOSCIENCES
LabCorp Specialty Testing Group

Top Five RAMs within Each Drug Class (N~7000 samples submitted for GS-Archive testing)

DRUG CLASS	MUTATION	% SAMPLES
NRTI	M184V	23.90
NRTI	M41L	15.46
NRTI	D67N	13.90
NRTI	K70R	13.26
NRTI	T215Y	13.24
NNRTI	K103N	17.17
NNRTI	Y181C	7.45
NNRTI	V108I	5.64
NNRTI	G190A	4.40
NNRTI	K101E	3.54
PI	L90M	11.14
PI	M46I	8.63
PI	V82A	6.51
PI	I54V	5.72
PI	G73S	5.59
INI	N155H	0.82
INI	G140S	0.52
INI	Q148H	0.39
INI	E138K	0.32
INI	S147G	0.27

Interrogating HIV DNA in the Setting of Antiretroviral Drug Suppression
13th European HIV & Hepatitis Meeting; June 05, 2015

NGS Comparative Study: Forum for Collaborative HIV Research

Sample Types	Sample Source
Well-characterized virus stocks (PR/RT/ENV)	SeraCare
Defined mixtures of PR PCR products (PI)	U North Carolina
Defined mixture of RT PCR products (NNRTI)	BC Center for Excellence

	Roche 454	MiSeQ	PacBio RS
CBER/FDA - Viswanath Ragupathy		X	
CFE/VC - Richard Harrigan		X	
Harvard/BWH - Jon Li		X	
Monogram - Chris Petropoulos		X	
Pacific Biosciences - Ellen Paxinos			X
Pacific Biosciences - Roche			X
Quest Diagnostics - Ron Kagan	GS JR		
Siemens - AJ Chmura		X	
UCSD - Davey Smith, 454	FLX		
UCSD - Davey Smith, RS			X
UCSF - Teri Liegler		X	
UNC - Shuntai Zhou		X	

manuscript in preparation

www.hivforum.org

Summary/Conclusions

- NGS platforms are replacing conventional Sanger sequencing
  - Operational efficiency (pan-virus, pan-target sequencing)
  - Cost effective (COG offset by automated data analysis)
  - Performance: sensitivity, flexibility, objective data analysis
  - *Currently used for all MGRM HCV DR testing and HIV GS-Archive; converting all other HIV DR assays*
- Easier to implement, train and maintain?
- Potential to use v-DNA in place of v-RNA (sample stability)
  - Analysis of the HIV DNA compartment must handle G to A hyper-mutation artifacts (e.g. D67N, M184I, G190S, D30N)
- HIV-1 *gag* and *env* gene sequencing is more challenging than *pol* gene sequencing due to high prevalence of insertion/deletions

Acknowledgements

- Monogram R&D
- Monogram PDO
- Monogram Bioinformatics
- Monogram Clinical Reference Laboratory
- UCSF/SFGH: SCOPE Cohort; Steven Deeks, Jeffrey Martin

- ***Special thanks to***:
  - Jonathan Toma
  - Jennifer Cook
  - Owen Solberg
  - Jeannette Whitcomb
  - Yuping Tan
  - Suqin Cai