Developing an Analysis Pipeline for the Regulatory Review of Next Generation Sequencing Data

Eric Donaldson, Ph.D.
Clinical Virology Reviewer
FDA/CDER/OND/OAP/Division of Antiviral Products

The FDA Science Board

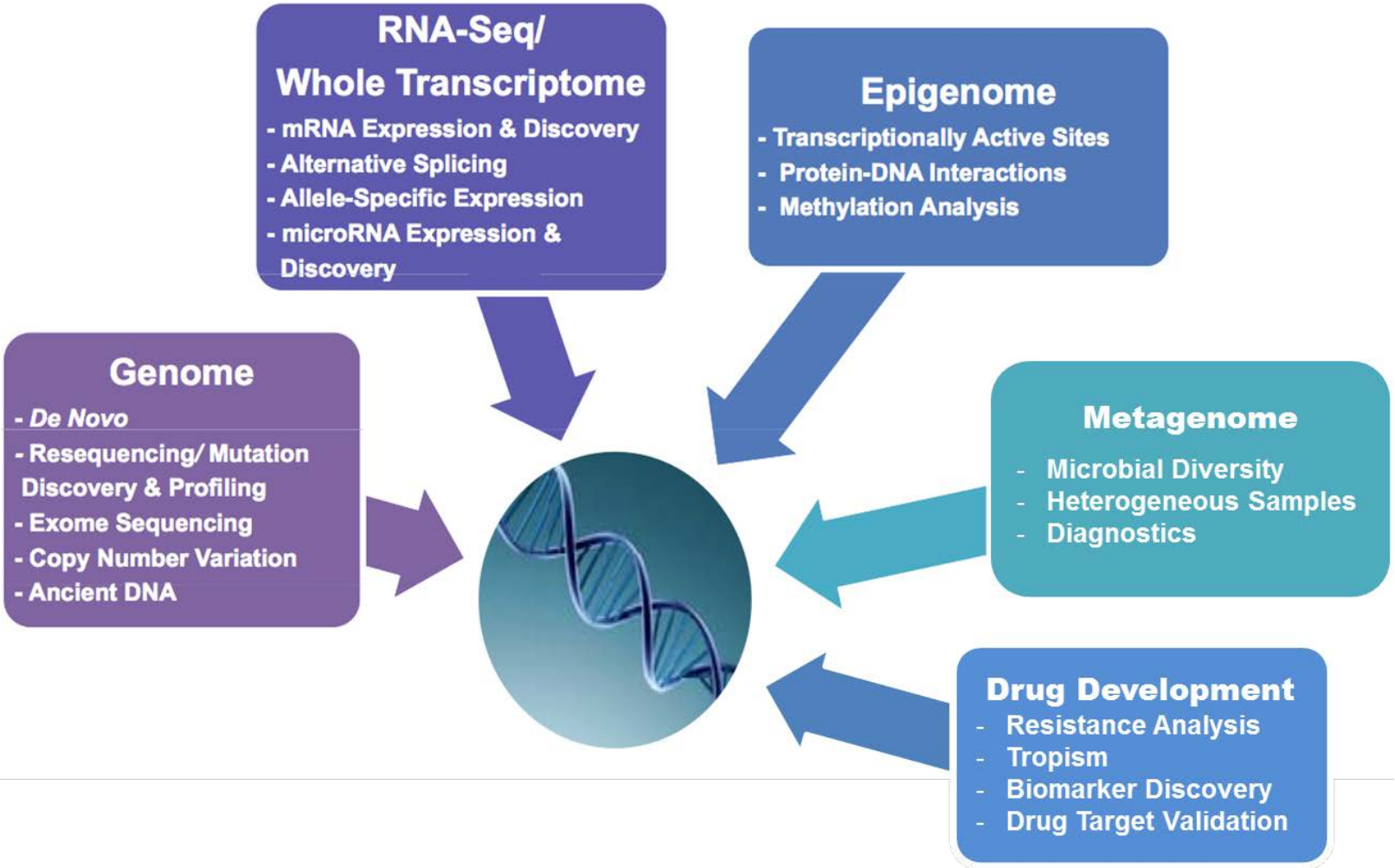
July 29, 2015

Disclaimer

My comments are an informal communication and represent my own best judgment. These comments do not represent FDA policy and do not bind or obligate FDA.

Cutting Edge Technology

Ultra High Throughput + Lower Cost = Broader Applications



NGS Activities In CDER

Current uses

- Viral resistance analysis
  - Drug durability issues
- Viral tropism analysis
  - Drug feasibility issues
- Detection of adventitious agents in viral isolates
- Metagenomics of microbial flora of the gut
- microRNA profiling in pancreatitis

Potential uses

- Personalized medicine
  - Individualized cancer treatment
- Biomarker discovery (GWAS)
- Identification of novel drug targets, e.g., oncology
- Identifying genetic susceptibilities to drug related adverse events

Viral Resistance as a Biomarker

- **Hepatitis C virus**
  - HCV GT1a with viral protease Q80K polymorphism does not respond to simeprevir
  - R155K or D168E in the HCV protease knocks out the whole class of HCV protease inhibitors impacting subsequent treatment options
  - L159F or C316N in the HCV polymerase at baseline may be predictive of sofosbuvir failure
- **Human immunodeficiency virus**
  - Baseline resistance profile of treatment-experienced subjects infected with HIV-1 who failed HAART
- **Influenza A virus**
  - H275Y substitution in the H1N1 influenza A virus neuraminidase confers oseltamivir resistance
- *Identifying resistance-associated substitutions and baseline resistance polymorphisms is essential for determining the safety and efficacy of a drug and for identifying subgroups of patients who will benefit the most from using it*

DAVP Review of Viral Resistance Data

- High level summary data in a format that shows polymorphic and treatment emergent amino acid substitutions (blank cells for insignificant changes)
- Tables listing each amino acid position with changes in decreasing frequency of occurrences in virologic failure subjects
- Information about assays and programs used for analyses and the parameter settings used for analysis
- **Raw data for independent FDA analysis**
  - Determine if treatment emergent amino acid substitutions can be correlated with treatment failure (emergence of resistance)
  - Determine if any baseline polymorphisms lead to reduced efficacy
  - Determine if resistance-associated substitutions confer cross resistance to other drugs against the virus
- **NGS data adds complexity to resistance analysis**

Importance of NGS

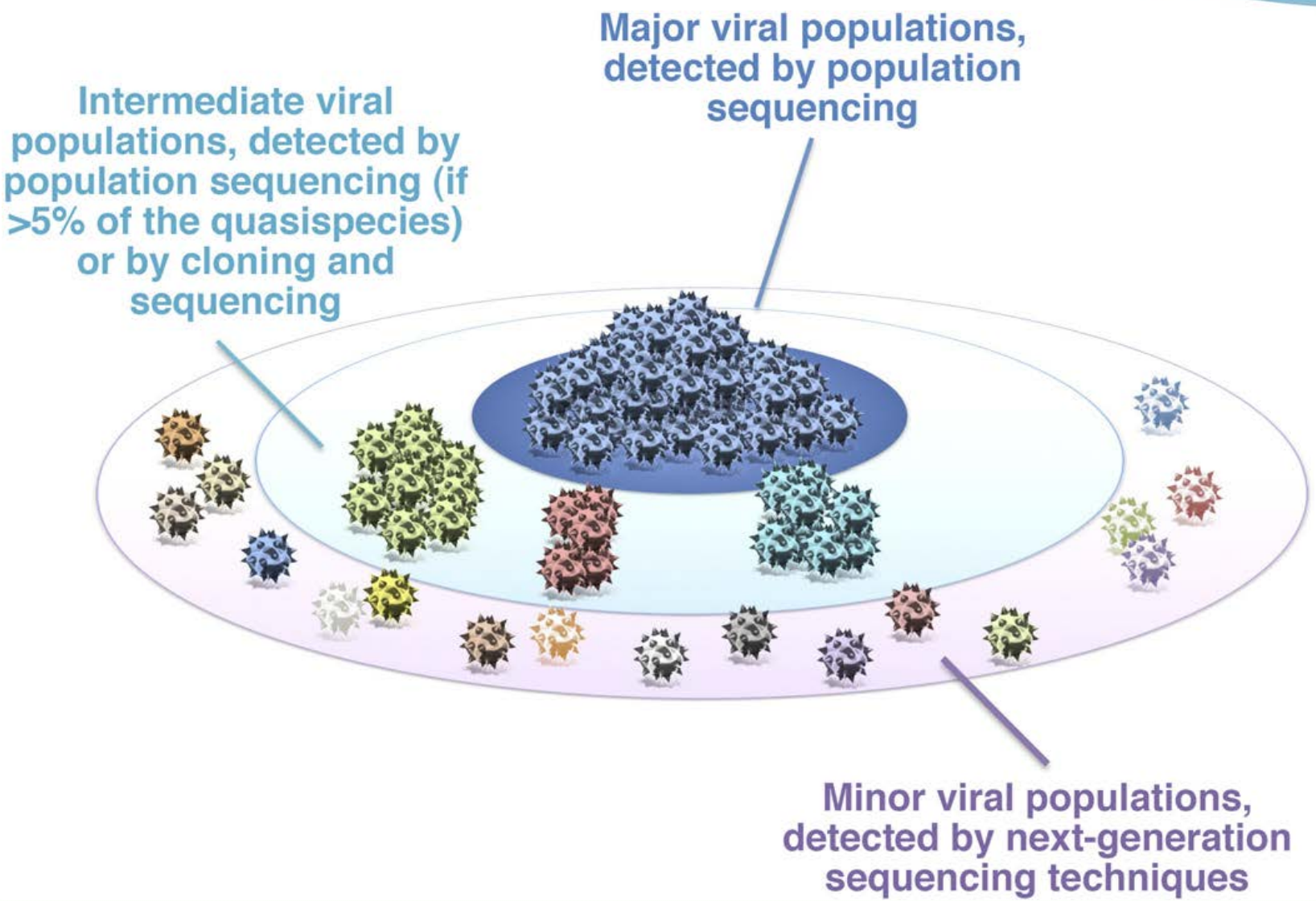
"Next generation sequencing has become a commodity." — *Vicky Schneider*

Why Sponsors Like NGS

Method	Single-molecule real-time sequencing (Pacific Bio)	Ion semiconductor (Ion Torrent sequencing)	Pyrosequencing (454)	Sequencing by synthesis (Illumina)	Sequencing by ligation (SOLiD sequencing)	Chain termination (Sanger sequencing)
Read length	5,000 bp average; maximum read length ~22,000 bases	200 bp	700 bp	50 to 250 bp	50+35 or 50+50 bp	400 to 900 bp
Accuracy	99.999% consensus accuracy; 87% single-read accuracy	98%	99.90%	98%	99.90%	99.90%
Reads per run	50,000 per SMRT cell, or ~400 megabases	up to 5 million	1 million	up to 3 billion	1.2 to 1.4 billion	N/A
Time per run	30 minutes to 2 hours	2 hours	24 hours	1 to 10 days, depending upon sequencer and specified read length	1 to 2 weeks	20 minutes to 3 hours
Cost per 1 million bp	$0.75-$1.50	$1	$10	$0.05 to $0.15	$0.13	$2,400
Advantages	Longest read length. Fast. Detects 4mC, 5mC, 6mA.	Less expensive equipment. Fast.	Long read size. Fast.	Potential for high sequence yield, depending upon sequencer model and desired application.	Low cost per base.	Long individual reads. Useful for many applications.
Disadvantages	Moderate throughput. Equipment can be very expensive.	Homopolymer errors.	Runs are expensive. Homopolymer errors.	Equipment can be very expensive.	Slower than other methods.	More expensive and impractical for larger sequencing projects.

Derived from BMC Genomics. 2012 Jul 24;13:341 and J Biomed Biotechnol. 2012:251364.

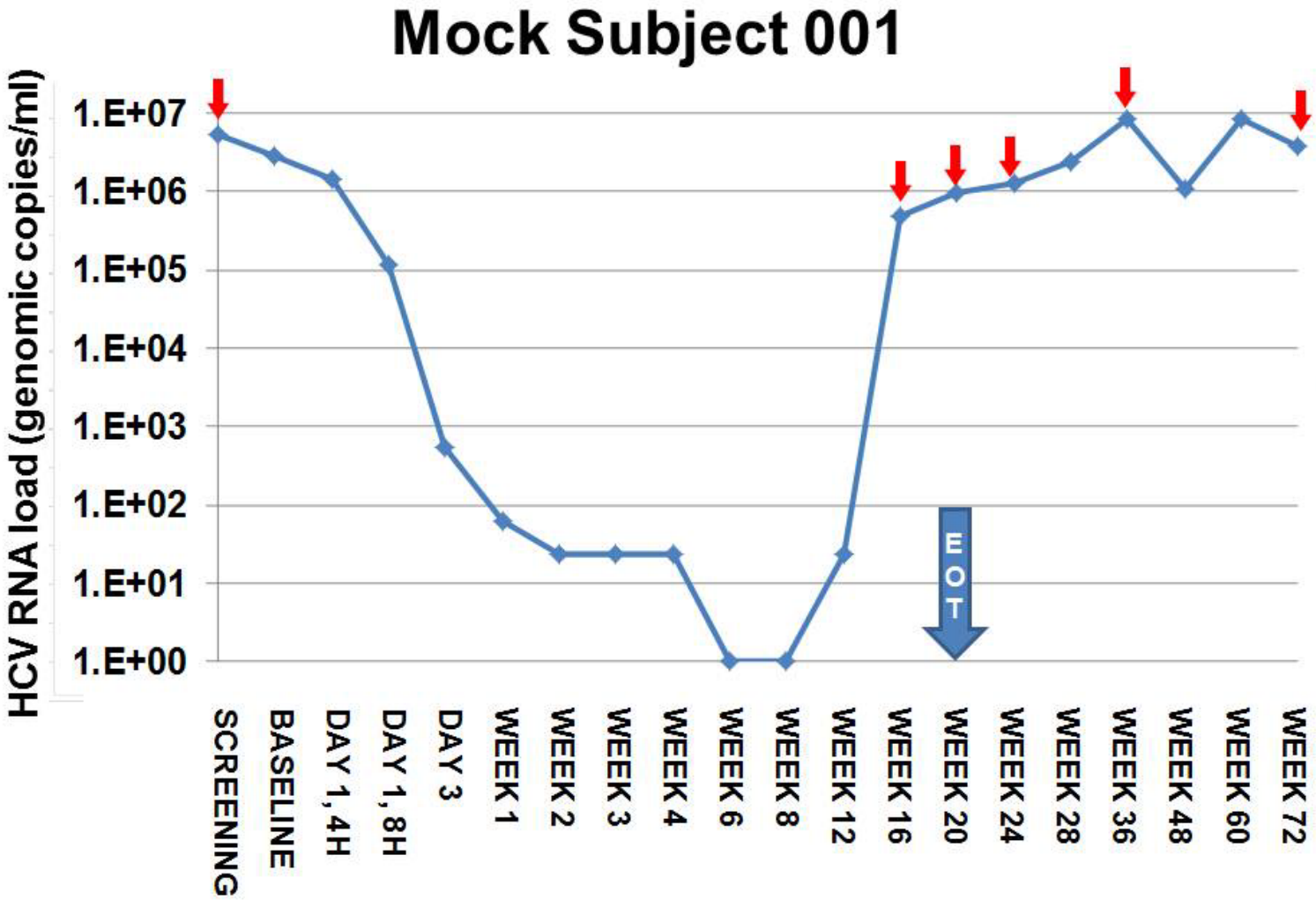
NGS Provides a Deeper Look



Chevaliez, et al. 2012. Gastroenterology, Vol. 142, No. 6.

Analysis of a Breakthrough

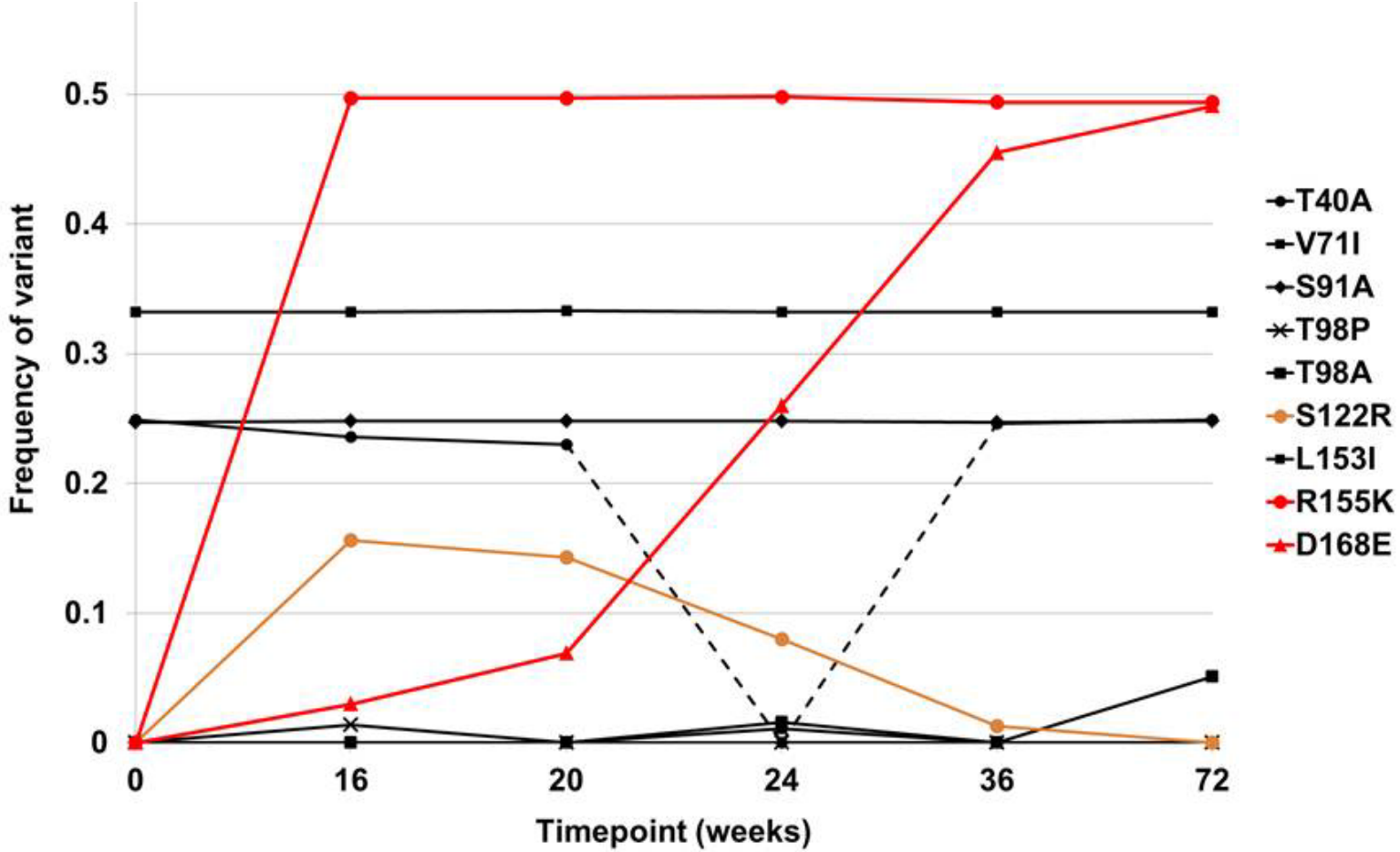
(Subject treated with an HCV protease inhibitor)



EOT = End of treatment

Observing Resistance Development

(Development of resistance to an HCV protease inhibitor)

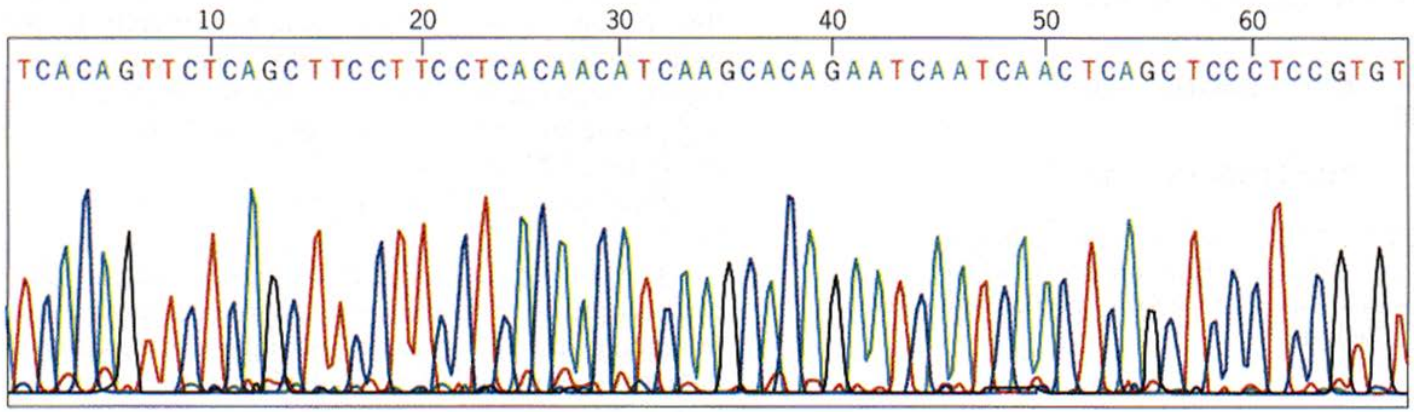


- *NGS allows observation of amino acid substitution frequencies over time*

The Technology Challenge

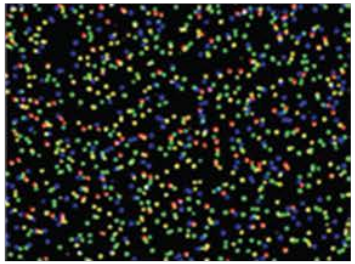
"There are no facts, only interpretations."
— *Friedrich Nietzsche*

NGS Data Complexities

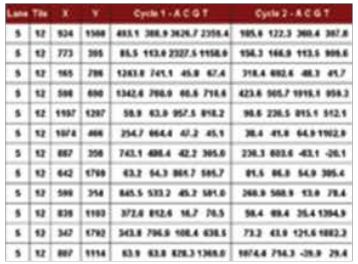


Sanger sequencing: one average viral sequence = one interpretation

Images

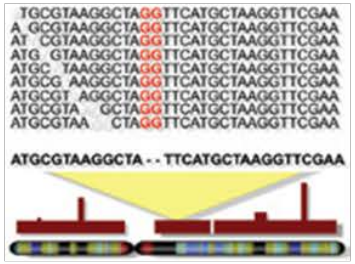


Intensities



Reads

Alignments



NGS: millions of viral sequences, hundreds of alignment algorithms, variable filtering criteria and assembly parameters, no standard analysis pipeline = many possible interpretations

Reference Mapping

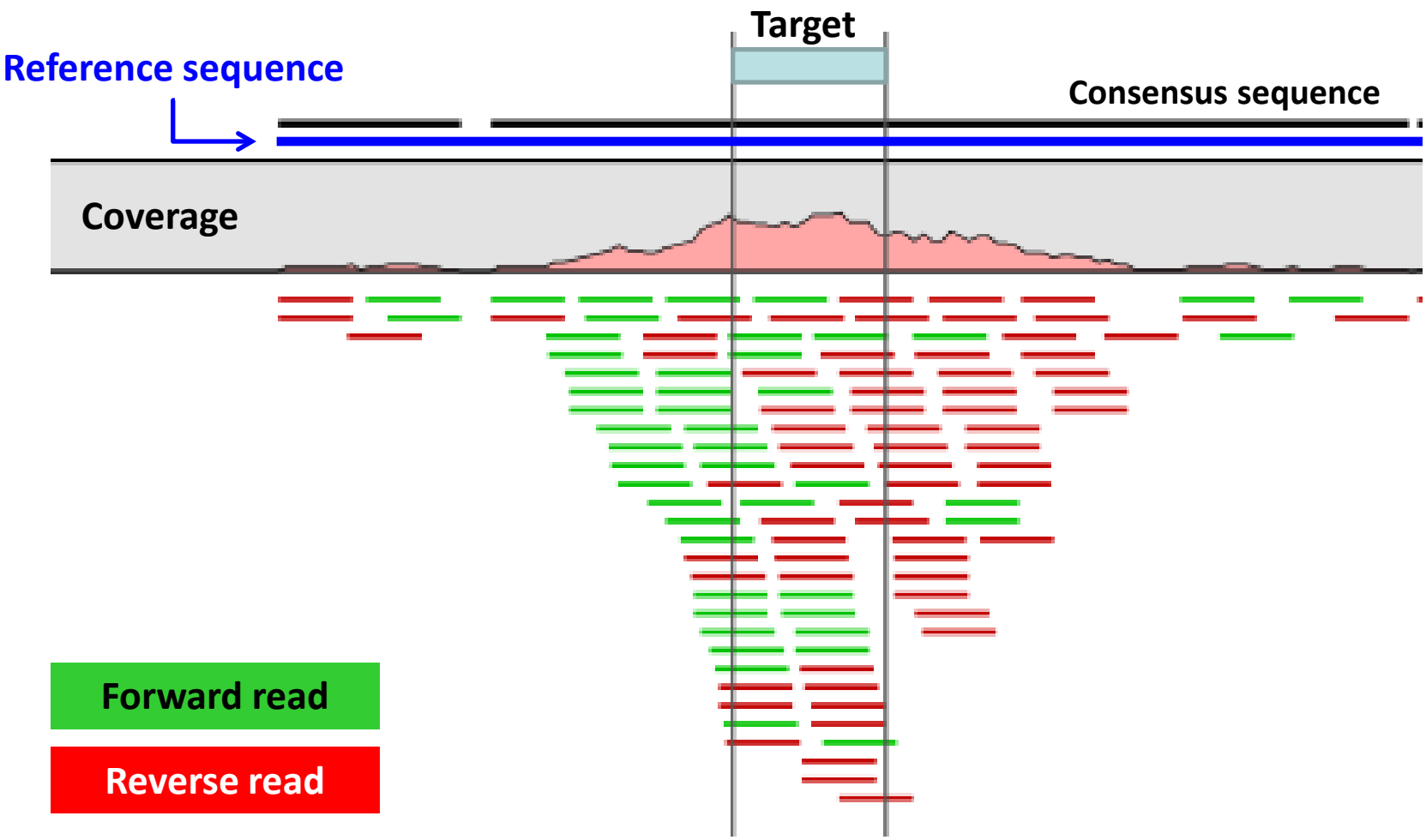
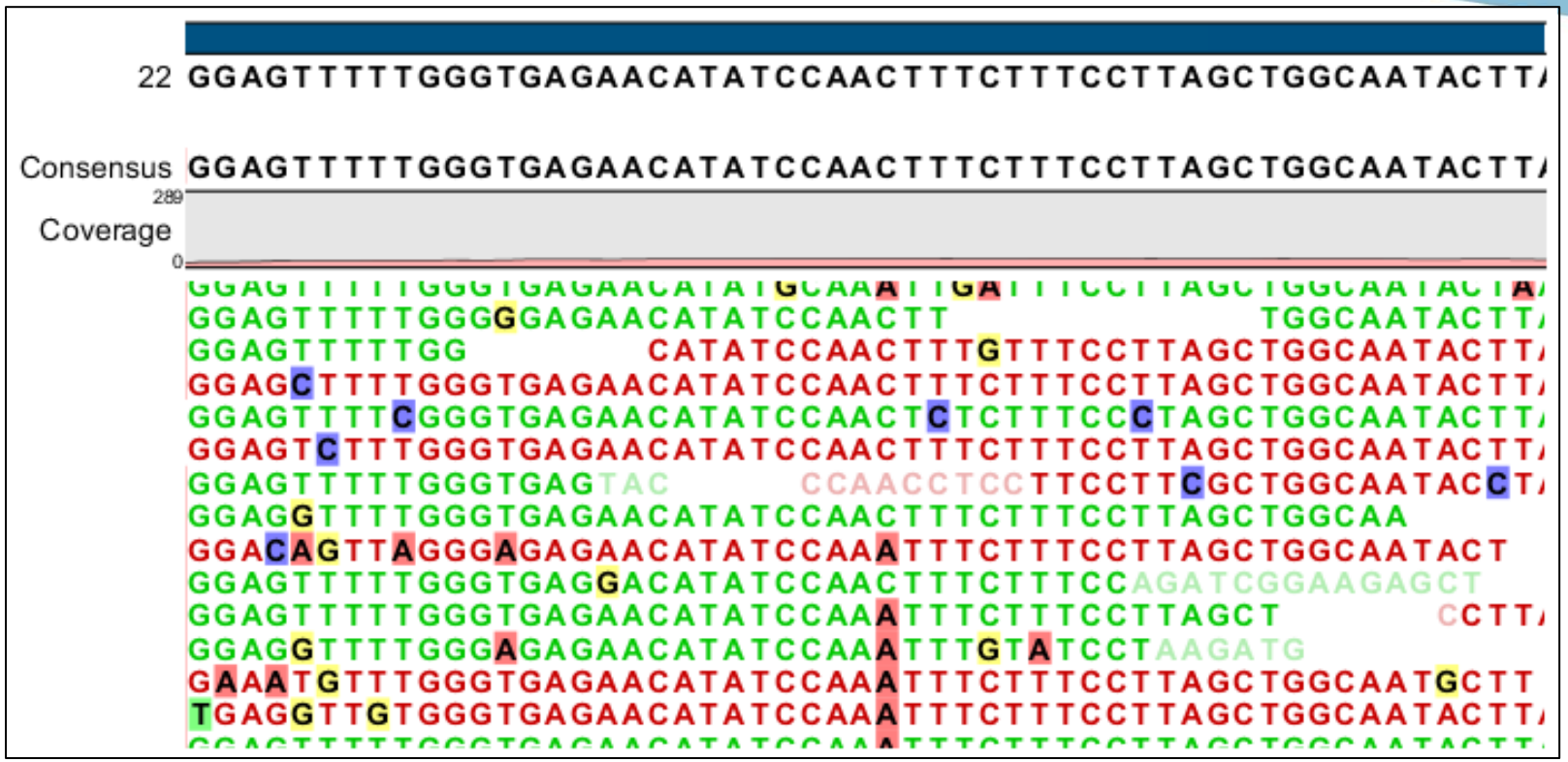


Figure from CLC Genomics Workbench Manual

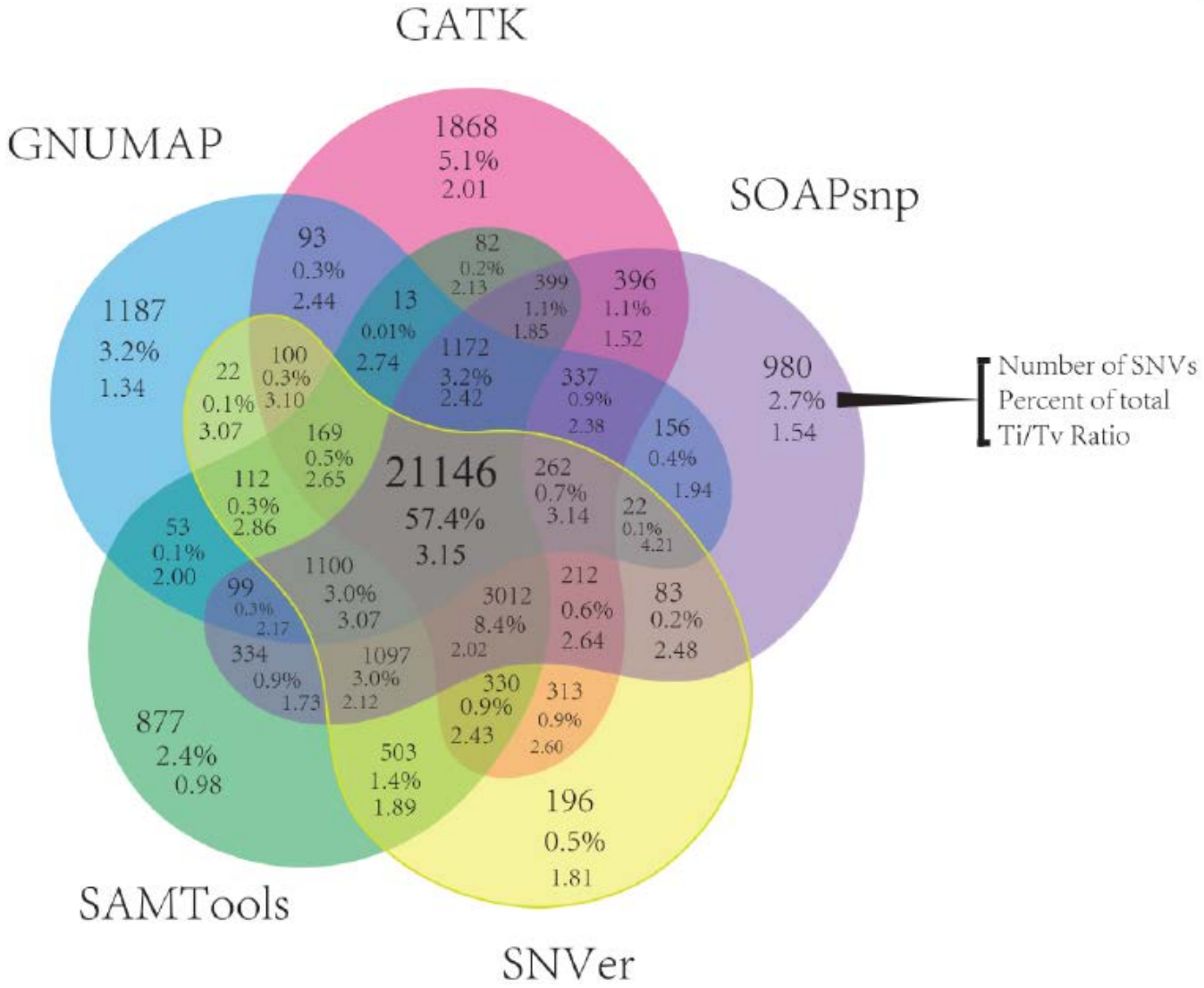
Variant Detection



- Different callers were designed for different purposes
- Very few standardized methods for SNP calling
  - Most methods: remove duplicates and filter bad reads
- Little overlap among the different callers

Figure from CLC Genomics Workbench Manual

Low Concordance SNP Detection



Review of NGS Data

"The truth is rarely pure and never simple."
— *Oscar Wilde*

The FDA Technology Challenge

- **Industry adoption of NGS technology outpaced development of FDA review technologies**
- **Two high profile NDAs Spring 2013 for the treatment of Hepatitis C virus**
  - Simeprevir (NS3/4A protease inhibitor) and Sofosbuvir (NS5B nucleoside analog polymerase inhibitor)
  - Both submitted NGS data for viral resistance
  - Both were discussed by Public Advisory Committees
  - Both NDAs reviews are complete and the drugs were approved by DAVP
- **At the time of submission, CDER had no capability to adequately review/analyze these data**
- **At the time of submission there was no "business process" for transferring and storing the data**

Data Analysis Challenges

- **Challenges:**
  - Very large data (~1TB per NDA)
  - No business process
  - No identified tools to support full review process
  - No assigned/designated IT resources/hardware

- **Division of Antiviral Products teamed up with CDER Computational Science Center (CSC) to develop a solution in collaboration with CDRH and CBER**
  - 8-10 weeks timeframe to solve all of these issues while on the review clock
  - Required significant resources from CSC to rapidly develop solution
  - Required support of CDRH supercomputer resources and CBER's HIVE

Regulatory Issues Pertaining to NGS Data for Drug Approval

- **Experimental design:** What exactly is being sequenced? Is the target amplified? Is it multiplexed? Are minor variants preserved? Are there experimental artifacts?
- **NGS data analysis:** How are the data being analyzed? Are the results robust? What programs and parameters are being used?
- **Regulatory review:** What data are necessary to make a regulatory decision? Are summary data from one analysis pipeline sufficient?
- **Data transfer and storage:** If raw data are requested, how are these data transferred to the agency? How are these data stored long term? Is encryption required? How is data integrity managed?
- **Data confidentiality:** How are we protecting proprietary information?

Review of Sofosbuvir Resistance Data

- Determine if any HCV baseline polymorphisms lead to reduced efficacy
- Determine if treatment emergent amino acid substitutions can be correlated with treatment failure (emergence of resistance)
- Determine if resistance-associated substitutions of SOF confer cross resistance to other HCV drugs

No Resistance Detected Claim

No Resistance Detected in Four Phase 3 Clinical Studies in HCV Genotypes 1-6 of Sofosbuvir + Ribavirin With or Without Peginterferon

Reported by Jules Levin, NATAP, www.natap.org
AASLD 2013 Nov 1-4 Wash DC

Evguenia S. Svarovskaia,1 Hadas S. Dvory-Sobol,1 Christy Hebner,1 Brian Doehle,1 Viktoria Gontcharova,1 Ross Martin,1 Edward J. Gane,2 Ira M. Jacobson,3 David R. Nelson,4 Eric Lawitz,5 B. Nebiyou Bekele,1 Diana M. Brainard,1 William T. Symonds,1 John G. McHutchison,1 Michael D. Miller,1 Hongmei Mo1 1Gilead Sciences Inc., Foster City, CA; 2University of Auckland, Auckland City Hospital, New Zealand; 3Weill Cornell Medical College, New York, NY; 4University of Florida, Gainesville; 5Texas Liver Institute, San Antonio, TX

- This abstract was presented at the American Association for the Study of Liver Diseases meeting in November of 2013
- This claim was based largely on the fact that S282T was a resistance-associated substitution identified in cell culture that rarely emerged in the clinical trials

Treatment-Emergent Substitutions in Phase 3 Trials

- **L159F was present in 13 subjects who failed**
  - Known resistance substitution of HCV nucleotide inhibitors (with L320F)
  - Emerged in 6 GT3a subjects, 2 GT1a subjects, 1 GT2b subject and present at baseline in 4 GT1b post-transplant relapsers
- **E341D emerged in 13 subjects who failed** ◄ **PCR artifact**
  - Emerged in GT3a relapsers
- **C316N/H/F present at baseline in 10 subjects who failed**
  - All 10 were infected with HCV GT1b
- **V321A emerged in 5 subjects who failed**
  - Emerged in GT3a relapsers
- **S282T or S282R emerged in 2 subjects who failed**
  - S282T: GT2b relapser (12-week SOF monotherapy)
  - S282R+L320F: GT1a non-responder (SOF/RBV)

***All highly conserved positions**

The Sponsor's Response in 2014

L159F and V321A Sofosbuvir Treatment-Emergent HCV NS5B Substitutions

Reported by Jules Levin
AASLD 2014 Boston Nov 8-11

Evguenia S. Svarovskaia,1 Hadas Dvory-Sobol,1 Brian Doehle,1 Edward Gane,2 Ira M. Jacobson,3 David R. Nelson,4 Eric Lawitz,5 Diana M. Brainard,1 John G. McHutchison,1 Michael D. Miller,1 Hongme Mo1 1Gilead Sciences, Inc., Foster City, CA; 2University of Auckland, Auckland, New Zealand; 3Weill Cornell Medical College, New York, NY; 4University of Florida, Gainesville, FL; 5Texas Liver Institute, University of Texas Health Science Center, San Antonio, TX

Conclusions

- Deep sequencing detected L159F in 15% and V321A in 5% of patients at virologic failure in SOF studies
  - Intensification of SOF treatment with Peg-IFN or LDV reduced the emergence of these variants
- Treatment-emergent L159F and V321A did not impact the retreatment outcome with SOF+RBV±PEG
  - L159F and V321A were no longer detectable or were not enriched at the time of retreatment failure
- At baseline, L159F was detected only in GT-1 patients and only associated with virologic failure in pre-transplant patients treated for short durations with SOF+RBV
- The clinical significance of these variants remains to be determined

Conclusions

- Analysis of raw NGS sequence data allowed DAVP to identify:
  - Low frequency treatment-emergent substitutions of unknown clinical significance that appear to be resistance-associated
  - Cross-contamination of clinical trial samples
  - RT-PCR artifacts that confounded interpretation of NGS resistance data
- NGS may prove useful in identifying minor populations of resistant viruses present at baseline that could lead to treatment failure if the wrong regimen is prescribed
- Cross-center collaborations were important for building the necessary analysis pipeline for NGS data (ongoing efforts)
- Access to NGS data and analysis software allowed CDER reviewers to make better informed decisions during review

Acknowledgments

CDER/OND/OAP/Division of Antiviral Products
Debra Birnkrant, M.D. – Division Director
Jules O'Rear, Ph.D. – Virology Team Leader
Lisa Naeger, Ph.D. – Senior Virology Reviewer
Damon Deming, Ph.D. – Virology Reviewer

OTS/Office of Computational Science
Lilliam Rosario, Ph.D. – Office Director
Charles Cooper, M.D.
David Epstein, PMP

CBER/HIVE Team
Carolyn Wilson, M.D. – Director
Vahan Simonyan, Ph.D.

Office of Anti-Microbial Products
Ed Cox, M.D. – Office Director
David Roeder, M.S. – ADRA

Office of the Chief Scientist
Ross Felice, M.D.

CDRH Scientific Computing Lab
Brian Fitzgerald, Ph.D. – Director
Mike Mikailov
Fu-Jyh Luo