High Throughput Sequencing Data Challenges in FDA Regulatory Review

Eric Donaldson, Ph.D.
Co-Chair, FDA Genomics Working Group
Clinical Virology Reviewer
FDA/CDER/OND/OAP/Division of Antiviral Products

HTS Computation Standards for Regulatory Science Workshop
March 16, 2017
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

RNA-Seq/Whole Transcriptome
- mRNA Expression & Discovery
- Alternative Splicing
- Allele-Specific Expression
- microRNA Expression & Discovery

Epigenome
- Transcriptionally Active Sites
- Protein-DNA Interactions
- Methylation Analysis

Genome
- De Novo
- Resequencing/Mutation Discovery & Profiling
- Exome Sequencing
- Copy Number Variation
- Ancient DNA

Metagenome
- Microbial Diversity
- Heterogeneous Samples
- Diagnostics

Drug Development
- Resistance Analysis
- Tropism
- Biomarker Discovery
- Drug Target Validation
### HTS Activities at FDA

#### CFSAN
- Identification of pathogens causing foodborne outbreaks
- Characterization of microbial contamination of the food supply
- Pathogen risk assessment

#### CVM
- Whole genome sequencing to monitor antibiotic resistance
- DNA barcoding for identification of foodborne pathogens
- Biomarker discovery of inflammation in food animals
- Metagenomics of animal feed

#### CBER
- Vaccine safety
- Identification of vaccine contaminants
- Validation of virus challenge stocks
- HTS for HIV genotyping, tropism and drug resistance

#### CDRH
- General oversight approaches for NGS-based tests – germline and somatic applications
- Enhanced microbial sequence quality
- Human and microbial reference material selection
- Development of universal databases of clinically relevant variants
- FDA-ARGOS database – microbial reference genomes for regulatory use
HTS Activities in CDER

• 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
• Personalized medicine
  – Individualized cancer treatment
• Biomarker discovery (GWAS)
• Identification of novel drug targets, e.g., oncology
• Identifying genetic susceptibilities to drug related adverse events
HTS Data Standards Issues

Because there are no universal HTS data and data analysis standards, several questions remain:

- What HTS analysis pipelines are being used?
- How are these pipelines being validated?
- What algorithms are being used for specific applications in a pipeline?
- Is there reproducibility between algorithms?
- What parameters are used by the sponsor?
- How are the parameters optimized?
- How will the analysis pipeline be evaluated for regulatory review?
- What information will be required to make a regulatory decision?
CDER Antiviral Drug Resistance

• HTS is used to assess a viral target at baseline (before drug is administered) and near time of failure (on drug treatment) to determine if amino acid changes can be identified that are resistant to the drug

• Resistance pathways are carefully described in the label of approved drugs

• Frequently include low frequency (n=2) resistance-associated substitutions

• The HTS analysis pipeline and optimization process is critical

• Independent assessment of the data has been required because of limitations of access to tools & pipeline information from the sponsor

• Review and verification of results is difficult because pipeline and parameters are often vague or not included or are treated as a "black-box"

• Our analysis pipeline contains manual data analysis that is not described in great detail
CLC and HIVE Analysis Pipeline

- **Upload read files and reference**
- **Map reads to reference:**
  - Detect indels and structural variants
  - Local realignment
- **Call variants:** Low frequency variant detector
- **Generate frequency tables & Identify substitutions:** Viral Mutation Comparator
- **Rename files:** Batch Renamer
- **Quality control:** MultiQC
- **Map reads to reference:** Hexagon alignment
- **Call variants:** Heptagon Profiling
- **Sponsor’s frequency table**
Regulatory Issues in HTS for Drug Approval

**Experimental design:** What exactly is being sequenced? What platform is used? 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? Are the results reproducible? What programs and parameters are used? Is the analysis pipeline publicly available?

**Regulatory review:** What data are necessary to make a regulatory decision? Are summary data from one analysis pipeline sufficient? How will the analysis pipeline be validated?

**Data transfer and storage:** How are raw NGS data transferred to the agency? How are these data stored long term? Is encryption required? How is data integrity managed? How are we protecting proprietary information?
HCV Antiviral Resistance BCO

See the Antiviral Resistance Detection Biocompute Object poster!
HTS analysis pipeline description & metadata
HCV Antiviral Resistance BCO

"parametric_domain": {
  "hexagon_minimum_coverage": "0.15",
  "hexagon_seed": "14",
  "hexagon_minimum_match_len": "66",
  "heptagon_freq_cutoff": "0.10",
  "heptagon_divergence_threshold_percent": "30"
},
"io_domain": {
  "reference_uri": [
  ],
  "input_uri_list": [
    "hive://nuc-read/514683",
    "hive://nuc-read/514682"
  ],
  "output_uri_list": [
    "hive://data/514769/dnaAccessionBased.csv",
    "hive://data/514801/SNP.csv"
  ]
},
"error_domain": [
  "false negative discovery < 0.10",
  "false positive discovery < 0.05"
]
HTS BCO Advantages

- **Transparency**: The BCO will allow the FDA to document the exact algorithms, versions, parameters, etc. that informed a regulatory decision. Sponsor’s will be able to submit a BCO that contains all of the information required to assess an HTS analysis pipeline and reproduce the HTS analysis results.

- **Reproducibility**: A BCO will contain all of the information necessary to reproduce the results of an analysis pipeline and will document software versions so that archived analyses can be properly evaluated and updated in the future.

- **Tractability**: Operating from a BCO can provide the foundation for understanding how an analysis pipeline is maintained & upgraded.

- **Interoperability**: A BCO from industry and from FDA should provide all of the information necessary for both parties to understand how a particular analysis pipeline works.
BCO Implementation Issues

• **The Data Standardization Challenge:** What are the essential components of an HTS analysis pipeline and in the absence of data standards, how does one conclude that the results are robust?

• **The BCO Buy-In Challenge:** A BCO provides specific details of an HTS analysis pipeline and contains information on how the components of the pipeline were applied to a specific problem (i.e., viral resistance). How do we ensure stakeholder engagement?

• **The BCO Pipeline Information Challenge:** A BCO contains information on the components of the pipeline used to analyze a specific HTS dataset; what information should be required for how the pipeline components were selected, validated, and tested?

• **The Regulatory Decision Challenge:** A BCO should provide all of the information required to assess the HTS analysis pipeline used by the sponsor and reproduce the analyses independently. What additional information will be necessary and what guidance?
Acknowledgements

**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
Pat Harrington, 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.
Vahan Simonyan, Ph.D.

**Office of Anti-Microbial Products**
Ed Cox, M.D. – Office Director

**Office of the Chief Scientist**
Khaled Bouri, Ph.D.
Shari Solomon, PM

**CDRH Scientific Computing Lab**
Stephen Whitney
Mike Mikailov
Fu-Jyh Luo

**George Washington University**
Raja Mazumder, Ph.D.
Charles Hadley King
Challenge 1

• **The Data Standardization Challenge:** HTS Data Standards are lacking, making it difficult to assess whether the results coming from one HTS analysis pipeline are robust, reproducible, and comparable to the results from the same data generated using a different analysis pipeline.

• What are the essential components of an HTS analysis pipeline and in the absence of data standards, how does one conclude that the results are robust?

• What are the minimal requirements for a BCO that would capture a robust HTS analysis pipeline?
Challenge 2

• **The BCO Buy-In Challenge:** A BCO should provide specific details of an HTS analysis pipeline and contain information on how the components of the pipeline were applied to a specific problem (i.e., analysis of resistance in subjects infected with GT1 HCV who were treated with a specific drug). It may or may not contain information on how the pipeline was validated, but it should include information for all components of the pipeline.

• What steps are necessary to ensure that different stakeholders are enthusiastic about creating/submitting a BCO?

• What proprietary issues may prevent pharmaceutical partners from buying into the BCO concept?

• How can regulatory agencies ensure that the proprietary rights of a BCO are protected?
Challenge 3

- **The BCO Pipeline Information Challenge:** A BCO should contain information on the components of the HTS analysis pipeline used to analyze a specific HTS dataset but it may or may not contain information on how the pipeline components were selected, validated, and tested. Moreover, there is likely disagreement regarding the best components available for specific uses.

- What type of information should a BCO contain regarding validation of the HTS analysis pipeline?

- What proprietary issues may prevent pharmaceutical partners from being willing to share this information?
Challenge 4

- **The Regulatory Decision Challenge**: A BCO should provide all of the information required to assess the HTS analysis pipeline used by the sponsor and reproduce the analyses independently.

- At what stage does a BCO become acceptable for regulatory review purposes? Will additional information still be required or will the BCO and summary data suffice?

- What guidance would be necessary for defining and agreeing upon a specific BCO for a specific HTS data application?

- How will this concept be initiated and implemented between different stakeholders?