Breakout on Sequencing

Participants: Scott Weiss, Howard Jacob, Peter Good, Jeff Schloss, Debbie Nickerson, Dan Roden, Adam Felsenfeld, Erwin Bottinger, Les Biesecker, Rex Chisholm, Nancy Cox, Scott Devine, David Ledbetter, Michael Murray, Gary Gibbons, Jennifer Morrissette, Adam Davis,

Goal: Change the practice of medicine to a point where a whole genome sequence (WGS) can be routinely ordered for a patient and to use it to improve their healthcare. What would community agree is reasonable set of data they want to accurately interpret from a genome and be able to act on it, so that it's useful for patient care? Before WGS will be accepted as a legitimate clinical test we first need to develop enough data to convince people that this needs to be in the clinic. What is a reasonable set?

Needs:

Set of standards, best practices for analysis (these may change with time)
- Alignment, representative of different major ethnic groups
- de novo
- Accuracy of calls for SNPs, genes, indels, CNVs, etc.

To know what we find with different search strategies

Well defined clinical phenotypes

Tiffany standards (equivalent to MIAMI standards for microarrays) for phenotypes and genome annotations.

Annotation strategies to meet the Tiffany standards and beyond

Layer different categories of data, SNP data, standard phenotypes from eMERGE and eMERGE-like EHRs onto genomic sequence.

What should a clinical data report look like? What data should be returned? Variants of unknown significance (VUS)—should it be defined differently, e.g. not defined.

Mission Critical: More and more labs are going to be rolling out WGS and exome sequencing. There is a critical need to define standards and meaningful clinical reports. Failure to do so will drive up costs. In this interim there will be pressure to run expensive clinical tests to follow-up WGS and Exome results.

Grand Vision: Sequencing 100,000 patients with detailed electronic medical records to build a comprehensive data set of variants, phenotype annotations, critical information about incidental findings. This would be analogous to the initial goal of sequencing the human genome. Need to complete some key pilot projects to set the specific needs and milestones for this project. A white paper could be developed this year to begin the dialogue. Should it be more grand—all children with all birth defects or "simply" all children?

Considerations: 1) Part of the question of getting WGS into routine testing is figuring out how to fund it – insurance, patients, research are the mix of fuming sources we use now. Sequencing cost will drop. If insurance or patient pays for the sequence it's free to the research. So use some research dollars now to build the infrastructure to enable it. 2) Pilot projects for disease specific questions would also fit into positioning the best approach for the Grand Vision.

Pilot Projects

Analytical Best Practices
Wetlab bake off
Improved reference set for clinical analysis
Establish minimum standards for genomic and clinical phenotyping data.
Work with NIST on developing "standard" genome types.
Create a central repository of WGS for clinical labs to compare against.

Pilot Projects

Analytical Best Practices

Goal: Develop a set standards for a software tool/analysis pipeline for clinical analysis

Needs: 10 whole genome sequences.
Set of references genomes to compare to.
Collection of defined clinical phenotypes.
6 genomes with known variants that cause disease
4 genomes with an unknown cause of disease—but having been analyzed.

Exercise: Interested groups would get the 10 genomes and reference genomes and conduct their own analysis. A bake off.

Meeting: To compare results write a lessons learned paper(s).

Deliverables:

A. Standard reference set for testing and bench marking future new tools
Coverage needed
B. Analytical guidelines
C. Best practices for validation of a clinical tool.
D. Potentially novel insights

Wetlab bake off

Goal: Compare sequencing strategies

Needs: 10 genomes consented—ideally the same as those for Analytical Best Practices

Exercise: Interested groups would use their in platform or strategy to analyze the 10 genomes.

Annual meeting: Compare new strategies or platforms

Deliverables

A. Coverage needs by platform
B. Best practices for sequence data generation

Improved reference set for clinical analysis

Goals: 1) Create a better reference set of genomes and phenotypes, 2) pilot the Grand vision

Needs: 500 genomes with detailed EHR.
100 from each of the five major continents
Subset with known rare variants, carriers, common but known variants.
Subset with extremes for a set of common clinical phenotypes e.g. blood glucose.
Potential sources of samples: eMERGE, eMERGE-like, George Church, other existing cohorts, dB Gap, CHOP--others.

Exercise: A grant mechanism?

Annual Meeting: To compare results and advance best practices.

Deliverables:

A. Standard reference set for testing and bench marking future new tools
Coverage needed
B. Analytical guidelines for different human populations
C. Best practices for validation of a clinical tool.
D. Potentially novel insights
E. Position for the Grand vision

4. **Establish minimum standards for genomic and clinical phenotyping data**.

Goal: Create minimum standards for data annotation for genomic and phenotyping data

Needs: Work group for genome annotation minimum standards.
Different databases, which ones?
Work group for phenotype annotation minimum standards.
What is a reference phenome?
What phenotypes travel together
Joint group to align genome and phenotype minimum standards?

Exercise: Meeting(s), and/or conference calls

Deliverables:

A. Position paper(s)
B. Potential variant catalog for clinical references

5. Work with NIST on developing "standard" genome types.

Goal: Work with NIST to be sure that standard DNA sample(s) are part of the pilot and/or Grand Vision.

6. Create a central repository of WGS for clinical labs to compare against.

Goal: Short-term solution—this group (or subset) agree to share genomes for comparisons.

Needs: Data share agreements MOU
Data security and data site(s)
Rules for data use

Long-term Pan-NIH