

From: [John West](#)
To: [Christian Haudenschild](#); [Richard Chen](#); [Mark Pratt](#); [Scott Kirk](#)
Subject: Prioritization of Ops Sequencing projects
Date: Friday, October 12, 2012 7:38:41 AM
Attachments: [Ops Projects Prioritization JW 11Oct2012.ppt](#)

All,

Following our discussion of the spreadsheet with the Ops projects, I have drafted a proposed set of priorities for the sequencing part of the list (attached). This still needs to be followed by a priority list on the Pipeline work. Comments are welcome. Let's review on Monday.

Thanks,

John

Current Ops Projects Prioritization Sequencing

Draft for discussion

John West

Oct 11, 2012

DNA Sequencing Project Priorities

(Details on following slides)

1. Getting the lab set up and basic capabilities demonstrated (mostly done)
2. Customer projects (Moyamoya & VA-setup)
3. R&D top priorities:
 - Fully modern WGS (sequencing completed)
 - Fully modern exomes (almost done)
 - Exome-Plus testing (needs a whole plan)
4. 1st CLIA Validation (24 CFTR exomes, replicates)
5. R&D Background projects (run when machine time is available)
 - NA12878 2x250⁺ & Mainstream diversity panel WES & WGS
6. CLIA Menu expansion to include WGS (w/ mainstream diversity panel ?)
7. R&D lower priorities:
 - Known pathogenic STR variant samples
 - Personalis-custom assays (pulsed field gel, phasing)
 - Accuracy dependence on sequencing operation parameters

Getting the lab set up and basic capabilities demonstrated

- Phi-X qualifying runs
 - Completed on both machines
- Whole genome sequencing
 - Good data from 8 lanes of NA12878 already completed
 - Further WGS lab optimization on other projects
- Exome capability demonstrated
 - Pullout looks OK on MiSeq
 - 1st run on HiSeq failed turn-around
 - 2nd run on HiSeq now, with 24 including CEPH family, Venter
- Establish data connection from machine output to HugeSeq pipeline
 - Casava ? done ?
- Upgrade HiSeq(s) to HS-2500 (Jan 2013 ?) & then validation run(s)
- Upgrade MiSeq to dual surface imaging (Date TBD) & then validation run
- Other ?

Customer Projects

- Moyamoya
 - 1st samples running now
- Getting set up for VA
 - Many items on R&D's current priority list
 - Hardware needs meeting this morning
 - Employee fingerprinting, etc
 - Christian to complete this list

R&D Top Priorities

- Fully modern WGS (sequencing completed)
 - v3 sequencing chemistry, PCR-free, gel-free, HiSeq run
 - Data is for A/B, MIE and single-genome error characterization
 - These also provide the WGS for WGS / exome / E+ comparisons
 - NA12878 we sequenced ourselves (8 lanes on HiSeq)
 - CEPH1463 family & Venter seq for us to 60x by ILMN
 - On pipeline runs, prioritize the replicates (NA12878 & Venter, for A/B comparisons)
 - Have we defined the Pipeline version for this ? NEMAR ?
- Fully modern exomes (sequencing almost done)
 - v3 sequencing chemistry, ILMN pullout, can't be PCR-free
 - Data is for coverage statistics & other “normal ranges” for QC report
 - These also provide the exomes for WGS / exome / E+ comparisons
 - Data can be used for A/B comparison of same sample & pullout sequenced twice, but does not include same sample with two separate library preps & pull outs (for that see CLIA validating CFTR exomes)
- Exome-Plus testing
 - Mark is developing a whole project plan & wants to review it
 - First round pullout kits are due in about 3 weeks

1st CLIA Validation

- 24 samples with independently characterized CFTR mutations
- Exome sequence with replicates of:
 - Library prep
 - Pullout
 - Sequencing
- Consider running each flow cell with an NA12878 exome as a control, esp for SV / CNV detection from exomes
- This data set is also of interest to R&D to understand exome variability & reproducibility
- For CLIA, the work must be performed by a CLS, start date : 10/22
- We will need to define pipeline version to run, etc

R&D Background Projects

Run when machine time is available

- NA12878 2x250+ full coverage WGS
 - Data is for our Gold-standard genome
 - Must be run on MiSeq to get these read lengths
 - May take 30-40 sequencing runs to accumulate enough coverage
 - Several runs have been completed so far
 - Test data set for approaches to alignment with reads >200 bases
 - Max 3 runs / week if MiSeq used full time
 - Shujun is experimenting with exact read lengths, likely to be asymmetrical (e.g. 275 / 250)
 - It may be helpful to have a range of insert lengths in this data set
 - Shujun may experiment with insert lengths 1-3 kbp
- Mainstream diversity panel WES & WGS
 - 7-8 mainstream but diverse ethnicities, 24 samples, DNA received from Coriell
 - This is to be an ethnically more diverse reference set for us, to parallel the work we have done on the European CEPH1463 pedigree
 - 24 exomes on HiSeq first (one flow cell)
 - Run with NA12878 as an exome control ?
 - Then 30-40x WGS as machine time is available (use for CLIA menu expansion to WGS ?)
 - These are also candidates for Exome+ testing
 - What Pipeline version ? NEMAR ?

CLIA Menu Expansion to Include WGS

- Overlap with WGS of mainstream diversity panel ?

R&D Lower Priorities

- Known pathogenic STR variant samples
 - These may be pulled forward as part of the Exome⁺ pullout testing, likely with long-read set, run on MiSeq
- Accuracy dependence on sequencing operation parameters
 - Some of this is being explored on the MiSeq for the Gold-Standard genome (NA12878 2x250⁺) and Exome⁺ pullouts
 - Read lengths (asymmetrical)
 - Insert lengths (how far can we go, particularly with pullouts ?)
 - Optimization for pullout of high GC set
- Personalis-custom assays (pulsed field gel, phasing)