

From: [Richard Chen](#)
To: [John West](#)
Cc: [Scott Kirk](#); [Christian Haudenschild](#); [MarkPratt](#)
Subject: Re: Sprint 4 pre planning - JW questions & input
Date: Friday, August 10, 2012 9:14:52 AM

let's talk sunday or monday john.

the disk system is supposed to begin installation today I believe.

On Fri, Aug 10, 2012 at 7:24 AM, John West <john.west@personalis.com> wrote:

- > The draft Sprint 4 list in Scott's email
- > below may include this (just in "short
- > hand") but it continues to be a top
- > accuracy priority to categorize the
- >>300k differences, when the same
- > genome is sequenced twice, by the
- > mechanisms which cause them, so we
- > can best prioritize potential solutions,
- > both bioinformatic and in the lab.
- >
- > Do we expect to have data back from
- > Illumina and Complete Genomics on
- > the CEPH1463 family and the Venter
- > genome ? If so, do we have a plan to
- > process and analyze them ?
- >
- > We should discuss lab priorities during
- > this sprint as well:
- >
- > CLIA development
- > Initial HiSeq runs (described in an email from Mark about 7/24)
- > Assay to resolve compressions etc using pulsed field electrophoresis
- > Development of Personalis-specific pull-out (carry over from Sprint 2)
- > Development of Personalis optimized methods for sequencing (in addition
- > to custom pullout)
- > Continued build-out of the lab facilities themselves
- >
- > As part of our exome effort, I wonder if
- > we should try processing an exome per
- > day in this sprint, again to get
- > experience and find what the normal
- > ranges are ? As I recall, there are a lot
- > of exomes in the SRA, some from
- > samples where we have already
- > processed the whole genome (good
- > to compare).
- >
- > Should we have an explicit goal to
- > get our new disk storage system
- > installed and working (storage
- > bottleneck removed) ?
- >
- > We can discuss more by email.
- > I should be back in the Bay area

> late Saturday (now on a flight Cuzco
> to Lima in Peru).
>
> Thanks,
>
> John West
>
> Cell (408) 836-5586
> jwest38261@aol.com
>
>
> On Aug 9, 2012, at 7:01 PM, Scott Kirk <scott.kirk@personalis.com> wrote:
>
> Hi,
>
> We're closing Sprint 3 tomorrow and will be planning Sprint 4 in the
> afternoon. The areas we're targeting are in the list below. It consists of
> the remaining work for prototype and will include some things carrying over
> from Sprint 3.
>
> I had a chance to go through a first pass with Rich before he left and have
> updated it somewhat to reflect where things are at currently.
>
> Irealize you're both out and we'll have to walk you through many of the
> results when you get back next week. If you see this before then let me
> know if you have any comments. It's going to be a stretch to finish what's
> listed below in a sprint so we may need some help on priorities to stay
> focused.
>
> Thanks,
> Scott
>
>
>
> Some comments on Sprint 3:
>
> We're currently on day 12 of the 25 genomes in 25 days and have been working
> through Java issues on the servers and what looks like some temperature
> issues. We're at 9 genomes analyzed and have some fixes in place as of
> today which could allow us to catch up a bit with more runs now in progress.
> Annotation launcher has gone to 1.0 and now 1.1.
> We have the variant statistics and at-a glance in place as well as a report
> package design to demonstrate.
> The tool updates are implemented and Hugo's working on testing them now, but
> the server issues have been slowing that down.
> The reports and madsen browning algorithm work remain on track
> We're getting infrastructure in place for propagation of accuracy metrics
> (likely won't finish by tomorrow though)
>
>
> Here's what's on the list for Sprint 4 (in rough order of priority)
>
> Exome support
> Integration and testing work - we've incurred quite a bit of technical debt
> and we should expect there to be some work in finishing integrating and
> testing things so they function properly as a "pipeline".
> Case/control dry run - Running a case/control "experiment" with a set of

- > samples producing all outputs and packaging them up.
- >
- > significant blocks in getting data - waiting for dbgap approvals. If dbGap
- > comes through we'll need exome support for the data to use it. Short of that
- > we could make something up from the samples we have. This would allow us to
- > exercise all fo the functionality, but may not be the most scientific study.
- >
- > Completion of 25 genomes in 25 days effort
- >
- > started: 7/27, finish: 8/21
- > Calculate ranges
- >
- > Any items left from sprint 3:
- >
- > tools testing
- > other sub-tasks from board
- >
- > Personalis Report Package
- >
- > integrate the various pieces of the prototype design (with feedback)
- > Report explanations
- > automation
- > populate with ranges from 25 genomes work
- >
- > A/B comparison updates (indels)
- > Long indel detection (deletions) support
- > VNTR support
- > Addt'l ethnic specific reference support needs
- > BAMQC updates (if any)
- > These need to be done, but are probably lower priority
- >
- > CG post aligned data as input (if necessary, it may not be with use of the
- > CG converter)
- > Implement integration design for LIMS (we really don't have a system on line
- > yet to integrate with)