

From: [Nan Leng](#)
To: [Hugo Lam](#)
Cc: [Shujun Luo](#); [John West](#); [Christian Haudenschild](#); [Mark Pratt](#); [Christian Haudenschild](#); [Richard Chen](#)
Subject: Re: MiSeq 2x250 run completed over the weekend
Date: Tuesday, August 28, 2012 10:19:14 AM

Hi All,

I downloaded the 2x250 dataset from basespace into here /local/science-ops-1a/data/miseq/A1NGW/A1NGW_S0_L001_R1_001.fastq.gz

I installed smalt which is the next generation aligner from Sanger after SSAHA2 (<http://www.sanger.ac.uk/resources/software/smalt/>).
time /home/nleng/bin/smalt-0.6.3/smalt_x86_64 map -i 800 -f samsoft -o map_1.sam -n 1 /science/data/index/smalt/hg19 A1NGW_S0_L001_R1_001.fastq
A1NGW_S0_L001_R2_001.fastq

The initial quick test show smalt works. I got the map.sam file within 40min using 40core which is slower than BWA-SW. It will take roughly 24hrs to align 1x of 2x250. I am still testing this brand new aligner and will collect more stats later on.

In the mean time, I will try to get MOSAIK and BFAST work.

Thanks,

Nan

On 08/27/2012 11:26 AM, Hugo Lam wrote:

Using BWA to align 1x of 2x300 took ~12 hrs. So a 40-core computer for a 40x 2x250 will still take 10-12 hrs to finish the alignment. If we run the alignment on all 6 R910s we have, we can cut it down to less than 2 hrs for the alignment. Nan will be helping out on testing couple more aligners and we will have a better assessment on the alignment and its quality once he has finished it.

On Mon, Aug 27, 2012 at 9:34 AM, Shujun Luo <shujun.luo@personalis.com> wrote:

We got 3.1G for this run, and 2.4G with Q>30. Read1 error rate is 0.9 and read2 error rate is 1.44

From: John West <john.west@personalis.com>

Date: Monday, August 27, 2012 7:20 AM

To: Microsoft Office User <shujun.luo@personalis.com>

Cc: Christian Haudenschild <christian.haudenschild@personalis.com>, Mark Pratt <mark.pratt@personalis.com>, Christian Haudenschild <chaudenschild@personalis.com>, Nan Leng <nan.leng@personalis.com>, Richard

Chen <richard.chen@personalis.com>, Hugo Lam <hugo.lam@personalis.com>

Subject: MiSeq 2x250 run completed over the weekend

All,

This was our first run with Illumina's newest sequencing biochemistry, rumored to have a new & better polymerase. They have released it on the MiSeq first and, as I understand it, plan a HiSeq version later this year, but only 2x150 bases there.

Shujun :

How many raw G bases resulted from this run ?

Is there a run # we can refer to it by ?

Nan :

Can you please download this from BaseSpace, convert the raw data (no GATK Q-score recalibration) to a SAM file and put it on a memory stick for me ? I will give you a memory stick for this when I get in.

Also, can you run the QC tool on this and circulate the result ? I am particularly interested in the GC bias.

Christian :

Do you know when our MiSeq will be upgraded to dual-surface imaging, which should double it's yield per run ?

Hugo :

What are your thoughts on alignment of this longer read data ?

Thanks,

John West

Cell [\(408\) 836-5586](tel:4088365586)
jwest38261@aol.com

On Aug 26, 2012, at 11:07 PM, Shujun Luo <shujun.luo@personalis.com> wrote:

Hi John,

We have the most updated chemistry for HiSeq, i.e., PCR free gel free sample prep, v3 cluster and sequencing chemistry.

I just finished one 2x250 miseq run with the new sequencing reagent, with the PCR free gel free Amy library. The seq matrix looks good. The data is

available at our basespace account. For some reason, there is no alignment data on basespace any more, and we will do a local alignment tomorrow.

Thanks,

Shujun

From: John West <john.west@personalis.com>
Date: Sunday, August 26, 2012 9:51 PM
To: Christian Haudenschild <christian.haudenschild@personalis.com>
Cc: Mark Pratt <mark.pratt@personalis.com>, Microsoft Office User <shujun.luo@personalis.com>, Christian Haudenschild <chaudenschild@personalis.com>
Subject: Re: Newest reagents for HiSeq operation

Hi Christian,

I think the SRA data we downloaded earlier this year might have been PCR-free but not gel-free ? Mark ?

It would be great to try a MiSeq 2x250 Amy genome. Do we have the best library for that ? Would that be PCR-free gel-free too?

Thanks,

John West

Cell [\(408\) 836-5586](tel:(408)836-5586)
jwest38261@aol.com

On Aug 26, 2012, at 9:35 PM, Christian Haudenschild <christian.haudenschild@personalis.com> wrote:

The data downloaded from illumina that is from january was V3 and no PCR I believe.
We are only running V3.
We have or should have received V2 miseq on Friday, so we could try a real 2X250 and see where it gets us, illumina upgraded the miseq software on thursday.
Christian.

From: John West <john.west@personalis.com>
Date: Sunday, August 26, 2012 8:57 PM

To: Mark Pratt <mark.pratt@personalis.com>, shujun luo
<shujun.luo@personalis.com>, Christian Haudenschild
<chaudenschild@personalis.com>
Subject: Newest reagents for HiSeq operation

Mark, Shujun, Christian,

I have been discussing with Mark how to use the results of our A/B comparisons to direct our development of accuracy-improving sample preps and bioinformatics. He has pointed out that the A/B comparisons we have so far are based on data from the Broad, using v2 Illumina sequencing chemistry. Now there is not only a v3 sequencing chemistry, but as I understand it, also a TruSeq PE Cluster kit v3-cBot-HS, which is supposed to significantly reduce GC-bias induced at high cluster density. On the Illumina web site, some of that seems to be fairly new, but it is a little hard to tell. For our in-house HiSeq runs, do we have these newest reagents ? It is pretty important for us to move our analysis forward from v2. If we do not have them yet, do we know when we will ? I am also not sure if we are following the gel-free PCR-free library prep protocols, or if there are issues with them ? They are also supposed to help a lot in reducing GC-bias.

Perhaps we can meet briefly Monday to discuss the next HiSeq run.

Thanks,

John

--
Hugo Lam, PhD
Project Manager, Computational Biology
Personalis Inc.
<http://www.personalis.com/>