New NGS sequencers, equipment, data delivery & storage options, and price changes: GSAF Update.

Scott Hunicke-Smith

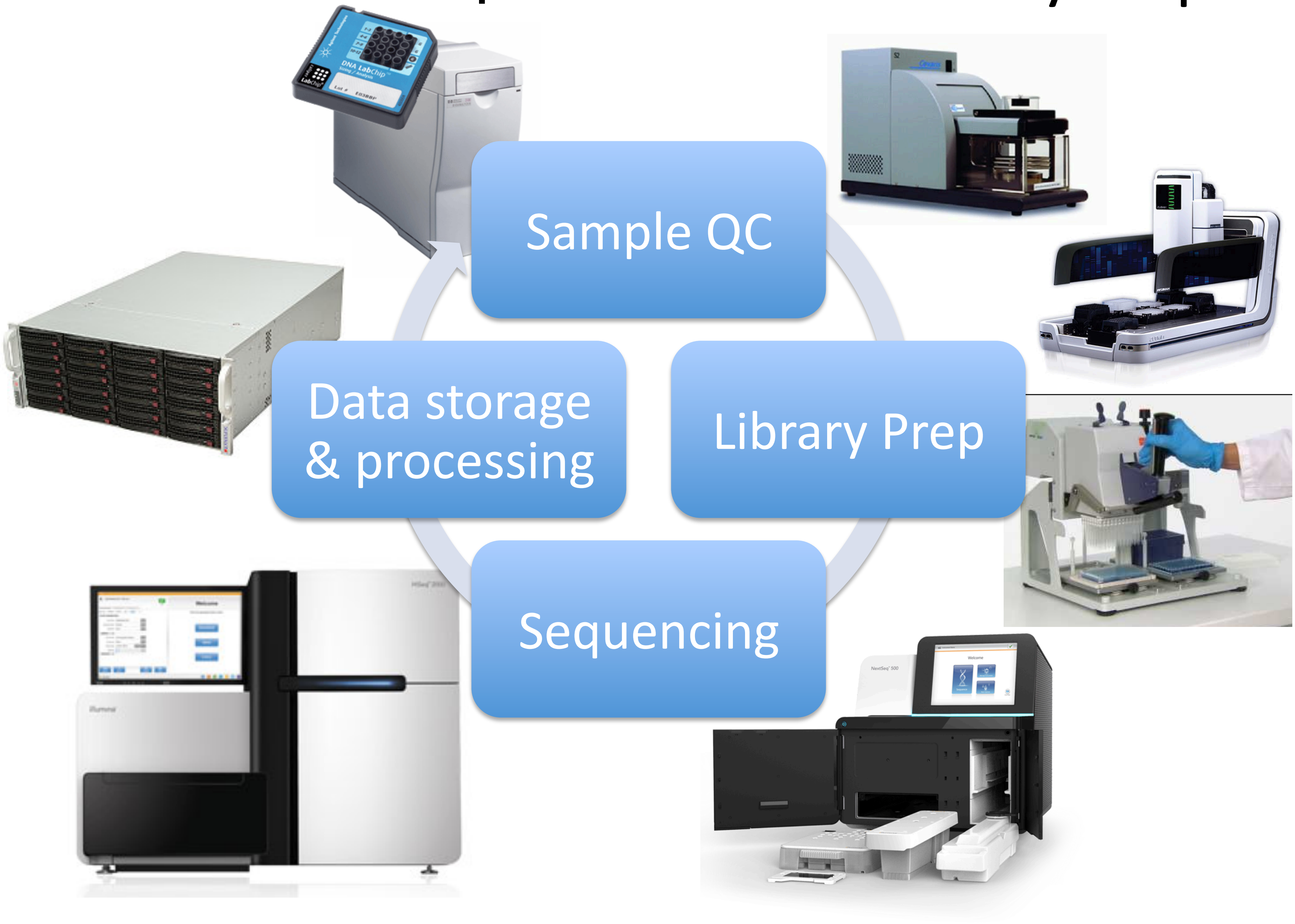
UT Byte Club

September 11, 2014

Outline

- Overview
- New user-accessible equipment
- NextSeq 500
- HiSeq Version 4 chemistry
- New library preps
- Data pods

Overview: Improvements in every step



Sample QC

Library Prep

Sequencing

Data storage & processing

GSAF FY2013-2014

- 810 jobs (vs 570 in 2012-2013)
- 7,480 samples (vs 4,820 in 2012-2013)
- 109 MiSeq runs
- 76 HiSeq runs
- 852 BioAnalyzer chips
- 273 qPCR plates
- 4 full-time staff (working really hard!)

QC: New user-accessible equipment

- Why? Improve quality, reduce bottlenecks!
- 4 instruments will be user-accessible (Oct):
  - Qubit Fluorimeter (Picogreen)
  - BioAnalyzer 2100 (DNA HS & RNA Pico)
  - Covaris S2 – Sonicator
  - Blue Pippin – size selection
- Calendaring via Google Calendar
- Self-service consumables sign-out
  - Qubit: Free
  - DNA HS: $50, RNA Pico: $30
  - Covaris: $2.50
  - Blue Pippin: $50/cartridge
- Training required first

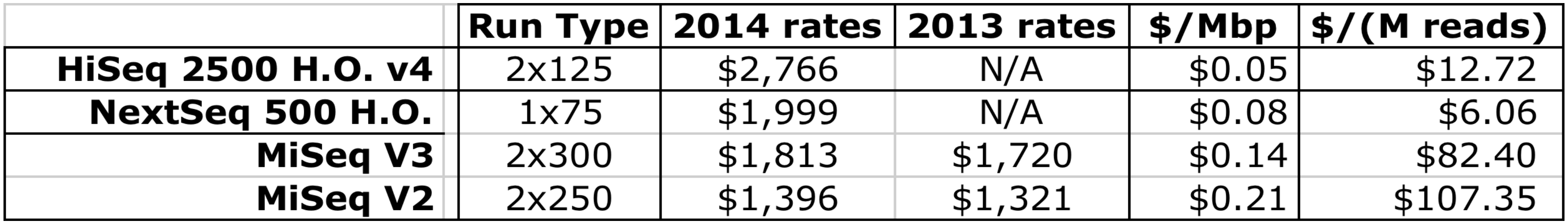
Illumina NextSeq 500 & HiSeq 2500 V4

- NextSeq 500:
  - SINGLE lane per run, 330M-400M reads
  - Cartridge-based like MiSeq with limited run types:
    - 1x75, 2x75, 2x150 – that's it!
  - Only 2 colors: "C" = red, "T" = green, "A" = red+green, "G" = nothing
  - ~10x larger clusters than HiSeq (i.e. a lot more amplification!)
  - Fast runs – 3 days max run
- HiSeq 2500 V4 chemistry & software "tweaks":
  - Max read length is now 2x125 (vs 2x100)
  - 33% increase in density (i.e. 218 M reads/lane vs 163)
  - 55% decrease in run time
  - Net: 3x improved throughput

Pricing

	Run Type	2014 rates	2013 rates	$/Mbp	$/(M reads)
HiSeq 2500 H.O. v3	1x50	N/A	$955	$0.12	$5.85
	1x100	N/A	$1,318	$0.08	$8.08
	2x50	N/A	$1,656	$0.10	$10.15
	2x100	N/A	$2,317	$0.07	$14.20
HiSeq 2500 H.O. v4	1x50	$1,013	N/A	$0.09	$4.66
	2x75	$2,091	N/A	$0.06	$9.61
	2x100	$2,428	N/A	$0.06	$11.16
	2x125	$2,766	N/A	$0.05	$12.72
NextSeq 500 H.O.	1x75	$1,999	N/A	$0.08	$6.06
	2x75	$3,329	N/A	$0.07	$10.09
	2x150	$5,025	N/A	$0.05	$15.23
MiSeq V3	2x300	$1,813	$1,720	$0.14	$82.40
MiSeq V2	2x250	$1,396	$1,321	$0.21	$107.35

Pricing



	Run Type	2014 rates	2013 rates	$/Mbp	$/(M reads)
HiSeq 2500 H.O. v4	2x125	$2,766	N/A	$0.05	$12.72
NextSeq 500 H.O.	1x75	$1,999	N/A	$0.08	$6.06
MiSeq V3	2x300	$1,813	$1,720	$0.14	$82.40
MiSeq V2	2x250	$1,396	$1,321	$0.21	$107.35

- Need fast turn-around on small jobs?
- Choose one of these run types!!
- (you will soon be able to pick “MiSeq 500 or 600 – whatever’s faster!)

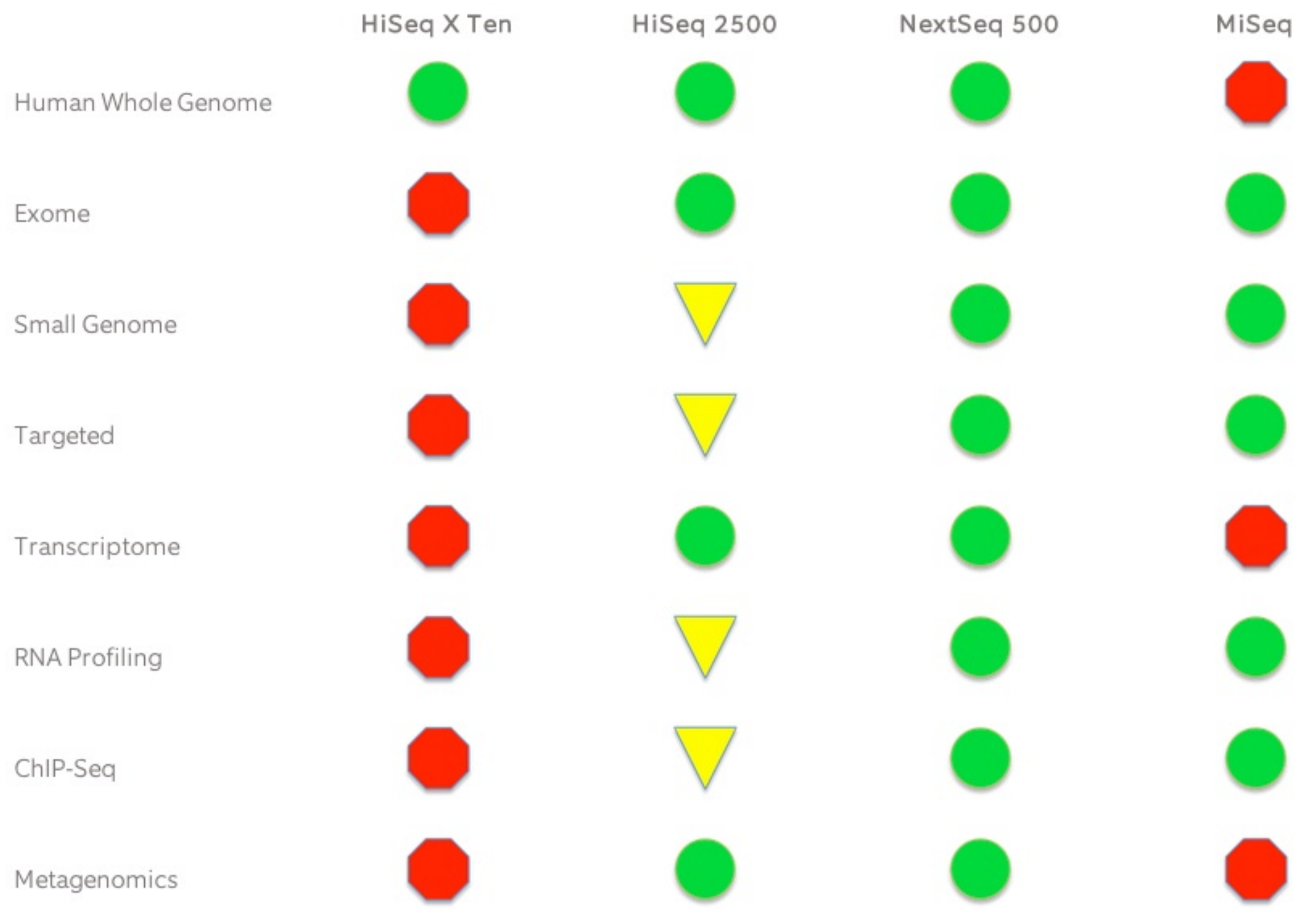
Summary Specifications

	HiSeq X Ten*	HiSeq 2500			NextSeq 500		MiSeq
		HT v4	HT v3	Rapid	High	Mid	
Total output	1.8 Tb	1 Tb	600 Gb	180 Gb	129 Gb	39 Gb	15 Gb
Run time	3 days	6 days	11 days	40 hrs	29 hrs	26 hrs	~65 hrs
Output/day	600 Gb	167 Gb	55 Gb	~110 Gb	~100 Gb	~36 Gb	~5.5 Gb
Read length	2 x 150	2x125	2x100	2x150	2x150	2x150	2 x 300
# of single reads	6B	4B	3B	600M	400M	130M	25M
Instrument price	$1M*	$740k	740k	$740k	$250k	$250k	$125k
Run price	~$12k	~$29k	~$26k	~$8k	$4k	?	~$1.4k
$/Gb	$7	$29	$43	$44	$33	?	$93

* Minimum purchase of 10 machines

From: Allseq.com - http://allseq.com/knowledgebank/sequencing-platforms/illumina

Suggested Applications by Machine



	HiSeq X Ten	HiSeq 2500	NextSeq 500	MiSeq
Human Whole Genome	🟢	🟢	🟢	🛑
Exome	🛑	🟢	🟢	🟢
Small Genome	🛑	▽	🟢	🟢
Targeted	🛑	▽	🟢	🟢
Transcriptome	🛑	🟢	🟢	🛑
RNA Profiling	🛑	▽	🟢	🟢
ChIP-Seq	🛑	▽	🟢	🟢
Metagenomics	🛑	🟢	🟢	🛑

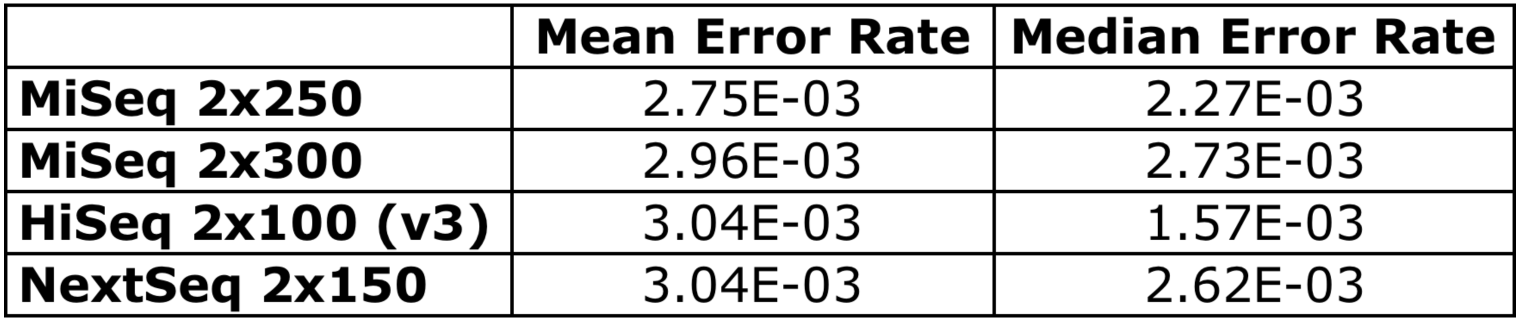
From: Allseq.com - http://allseq.com/knowledgebank/sequencing-platforms/illumina

Minimum Purchase Requirements

- NextSeq is ONE LANE – so barcodes matter!
- If your purchasing "by read" AND submitting finished libraries there are minimum purchase requirements:
  - Minimum purchase on NextSeq is the larger of:
    - 40 M reads per job
    - 20 M reads per sample
    - ASSUMING no BC overlap (the GSAF making your libraries will help here – we have lots of BC's)
  - Min. reads per sample on HiSeq V4: 20M

NextSeq vs MiSeq vs HiSeq

- Data quality from NextSeq OVERALL looks excellent…
- Some suspicion that additional amplification during clustering will differentially skew population of sequences detected based on size EVEN MORE than HiSeq
- Error rates vs. PhiX control:



	Mean Error Rate	**Median Error Rate**
MiSeq 2x250	2.75E-03	2.27E-03
MiSeq 2x300	2.96E-03	2.73E-03
HiSeq 2x100 (v3)	3.04E-03	1.57E-03
NextSeq 2x150	3.04E-03	2.62E-03

New Library Preps

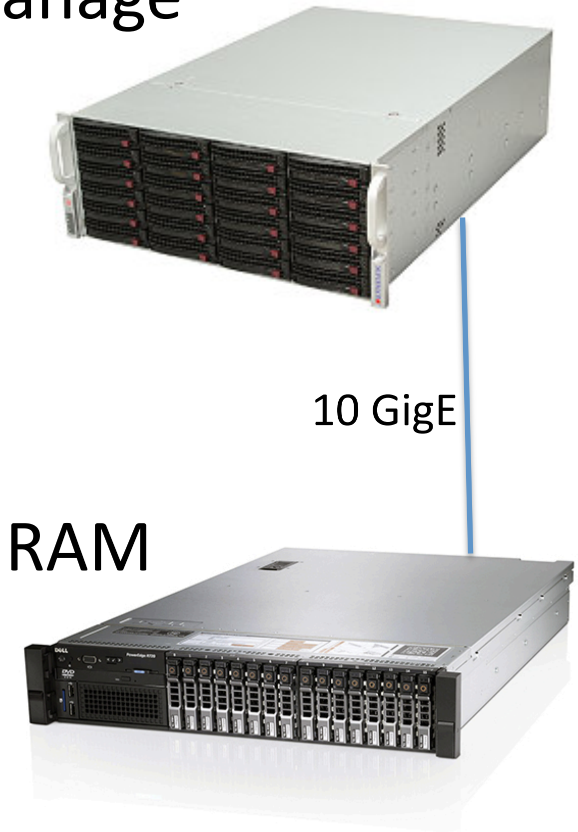
- New plate-based RNA & DNA library preps
  - RNA: $135/sample – including poly-A
  - DNA: $62/sample
  - Both include barcodes & QC

- Metagenomic/amplicon preps
  - Nested PCR design; 400 barcodes available
  - $21.25/sample for > 176 samples
- Genotyping by sequencing (GBS) preps
  - RAD-based; hundreds of barcodes available
  - $25/sample
  - “Development services” available to help design assay

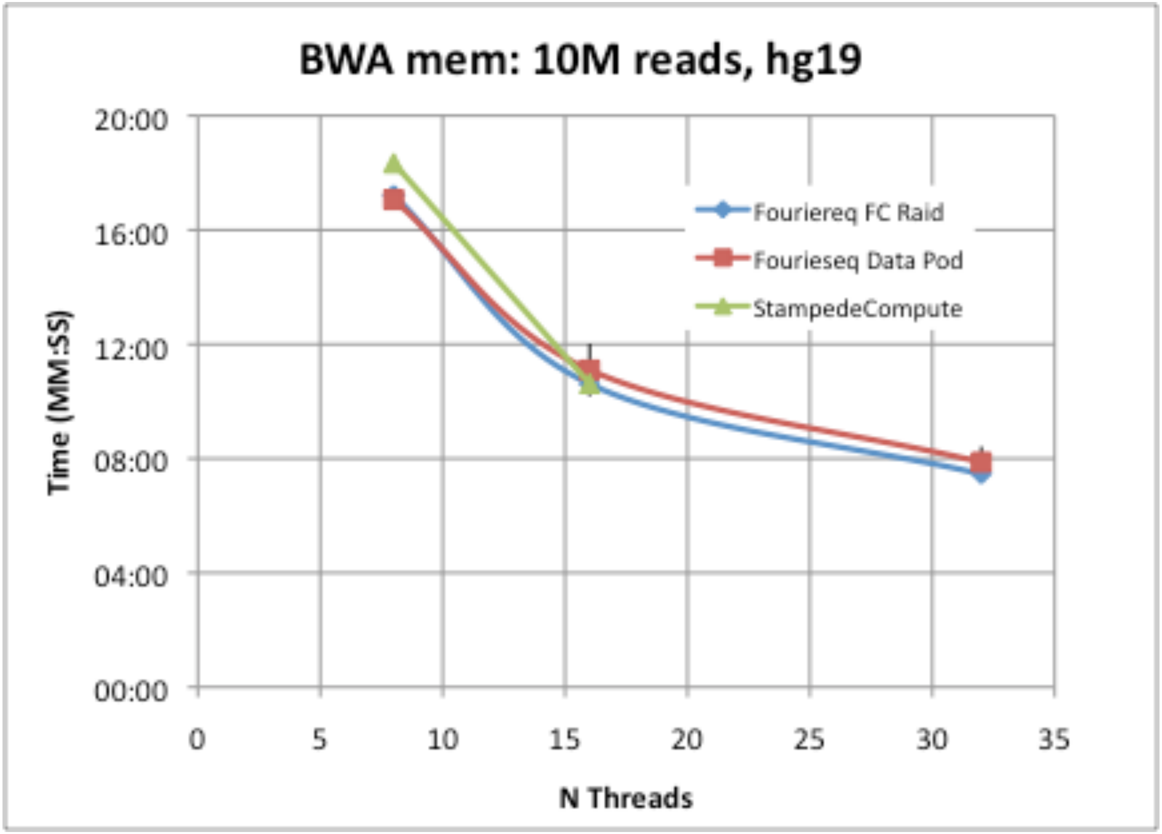
New Data Storage Options (coming soon)

- Home-built 96 TB data storage pods
  - > 80 TB formatted, @ < $90/TB TOTAL
  - ZFS file system – robust, easy to manage
  - Simplifies backup & recovery
  - Dual 10 GigE ports

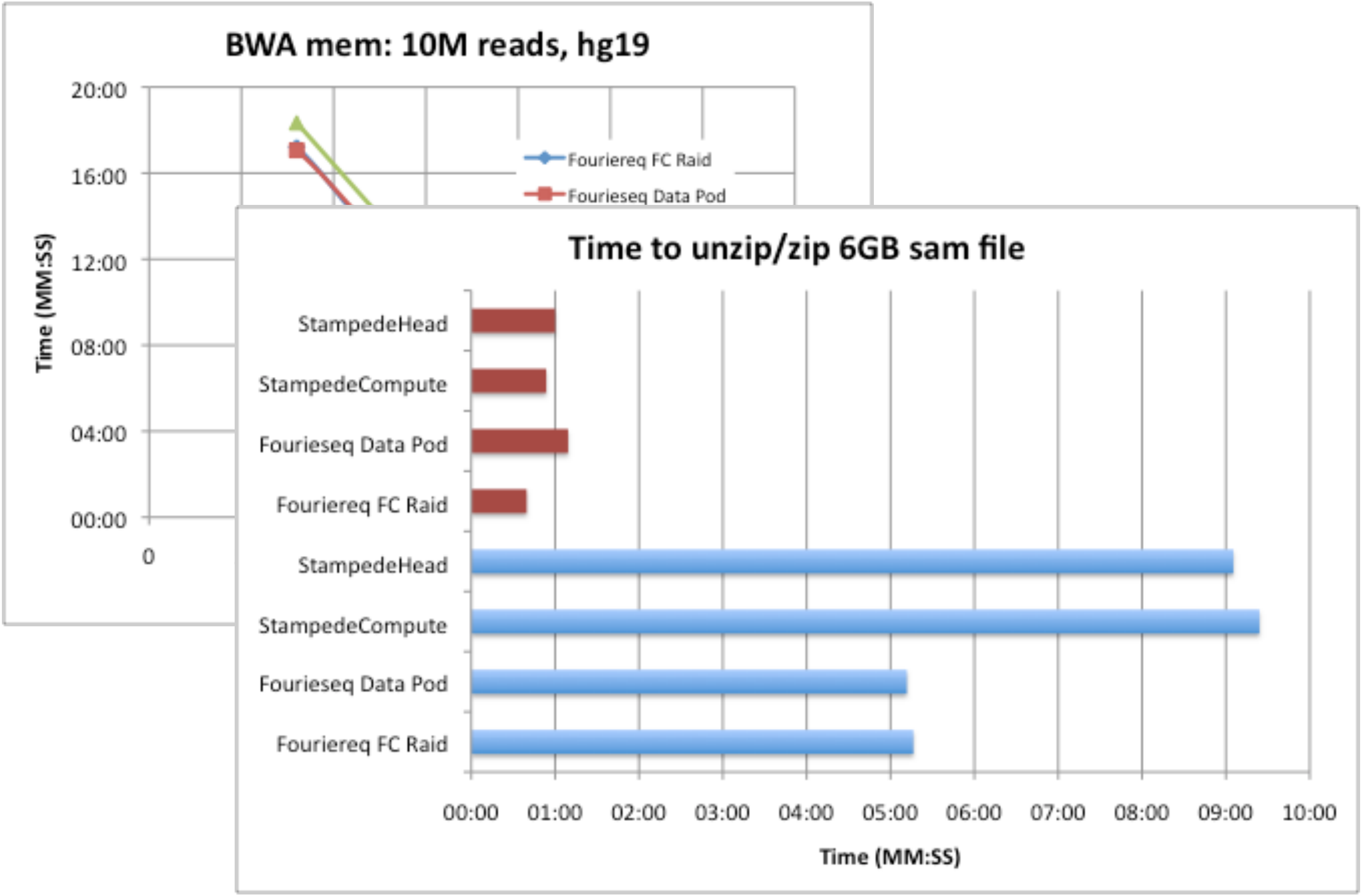
- New Fourierseq
  - 32 (hyperthreaded) cores, 192 GB RAM
  - 10 GigE connection to data pod



Data Pod/Fourierseq Performance



Data Pod/Fourierseq Performance



BWA mem: 10M reads, hg19

Time to unzip/zip 6GB sam file

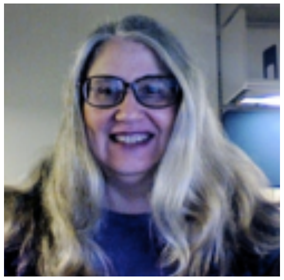
Other interesting software tidbits

- Software stack, esp. for python & R libraries
- Long-running jobs

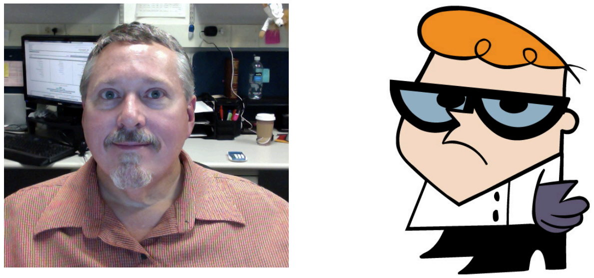
Credits…

> “Without continual growth and progress, such words as improvement, achievement, and success have no meaning”
>
> -Ben Franklin

Hans Hofmann/CCBB



Anna Battenhouse



Cecil Harkey & Charlie Gonzales

The GSAF: Jessica Podnar, Matt Barnette, Heather Diedrick, Terry Heckman, Gabby Huerta, Mani Singh