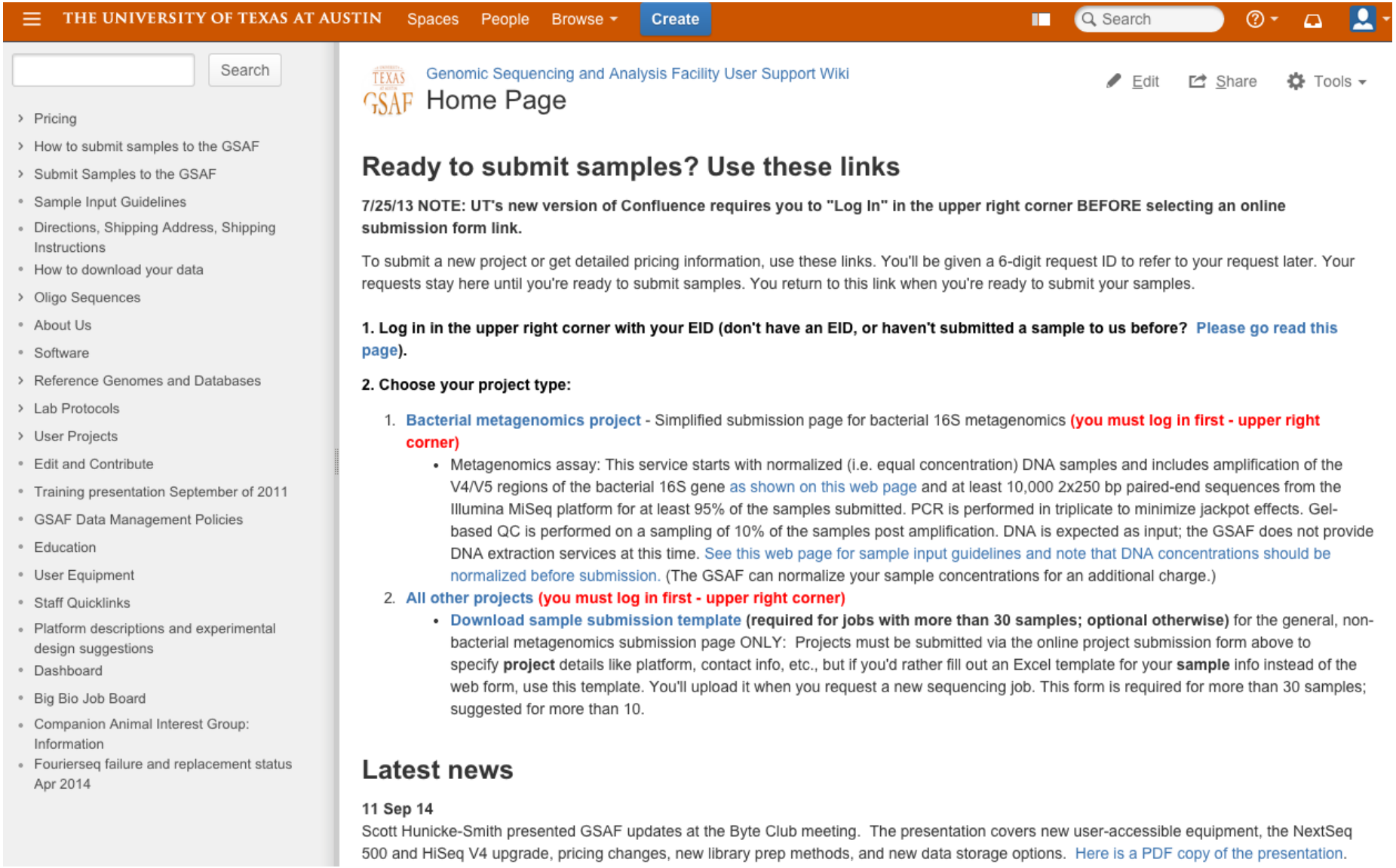
An Introduction to NGS at UT: The UT GSAF

Scott Hunicke-Smith

Updated: 9/23/14

Where to find more information

gsaf.cssb.utexas.edu



What the GSAF does

- Sample QC
  - By Fluorimetry/Qubit and BioAnalyzer
  - Self-service and full-service
- Library Preparation
  - Construction of NGS sequencing libraries from DNA and RNA of all sorts
- Sequencing
  - Final sequencing library quantification
  - Operation of the Illumina HiSeq 2500, NextSeq, and MiSeq systems
- Advise and consult
  - On experimental designs, lab methods, and cost

What the GSAF does

- Sample QC
  - By Fluorimetry/Qubit and BioAnalyzer
  - Self-service and full-service
- Library Preparation
  - Construction of NGS sequencing libraries from DNA and RNA of all sorts
- Sequencing
  - Final sequencing library quantitation
  - Operation of the Illumina HiSeq 2500, NextSeq, and MiSeq systems
- Advise and consult
  - On experimental designs, lab methods, and cost

850 BioAnalyzer runs/year
273 384-well qPCR plates/year

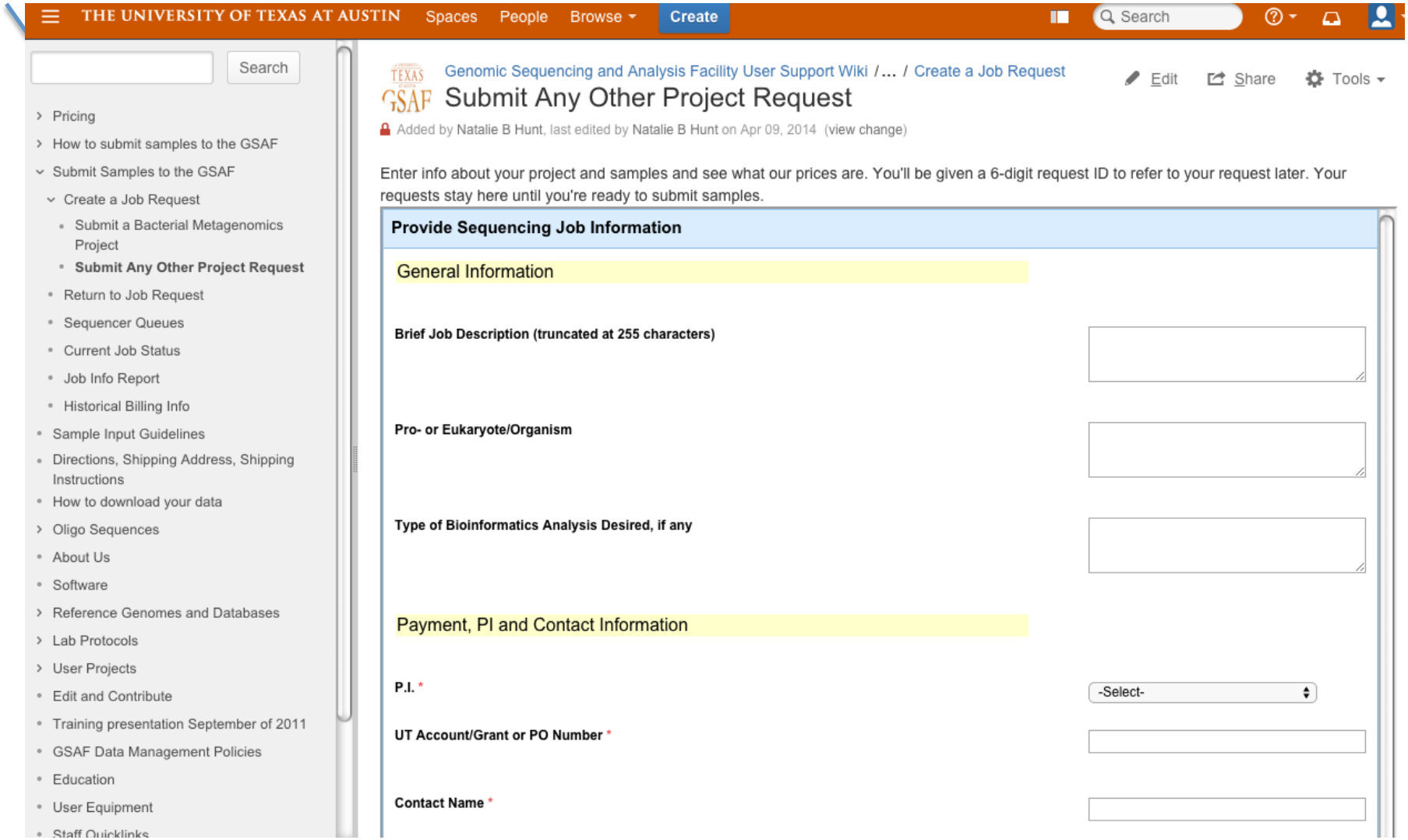
7,500 samples/year,
190 sequencer runs/year

800 jobs (projects)/year

What the GSAF does not do

- Data analysis
  - See the CSSB/CCBB Bioinformatics Consulting Group or
  - Investigate data analysis training available at UT (both via CCBB)
- Design your experiment
  - You, the customer, are responsible for your experimental design including the detailed choices of library preparation, sequencer run type, etc.
  - Illumina has some excellent tools and training videos that can assist when planning out projects, please visit http://support.illumina.com/ there is also an excellent resource for calculating coverage at http://support.illumina.com/downloads/sequencing_coverage_calculator.html
    - The coverage calculator will be extremely helpful when submitting your samples for sequencing and you need to determine the number of reads to request for each sample

How do I start a project with the GSAF?



How NGS sequencers work

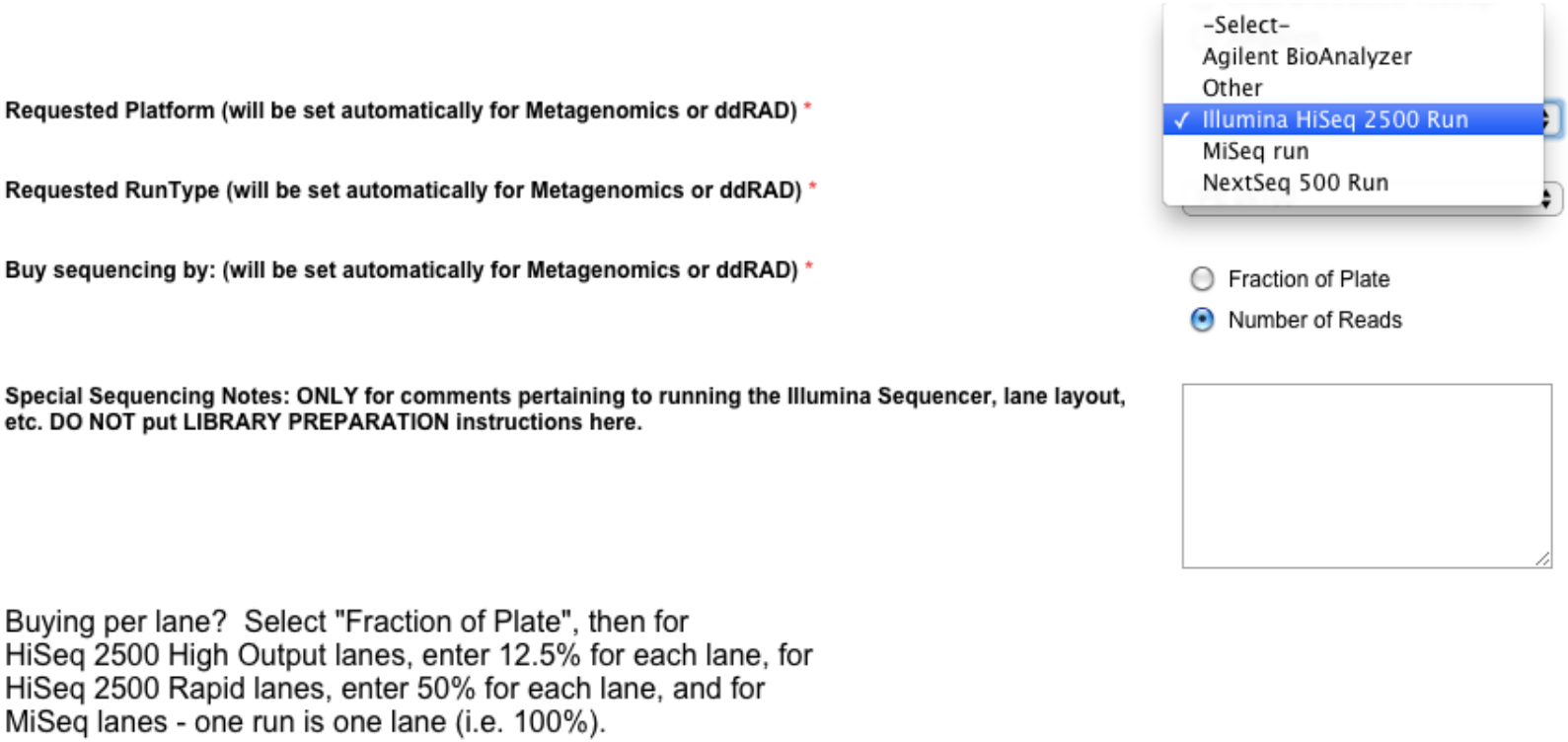
- Please see this YouTube video for a great illustration:
  - http://www.youtube.com/watch?v=77r5p8IBwJk
- Important practical conclusions:
  - Sequencing library DNA must be properly formed and suited to your question
  - The quality and quantity of sequencing library is essential

Sequencer run types & price

	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

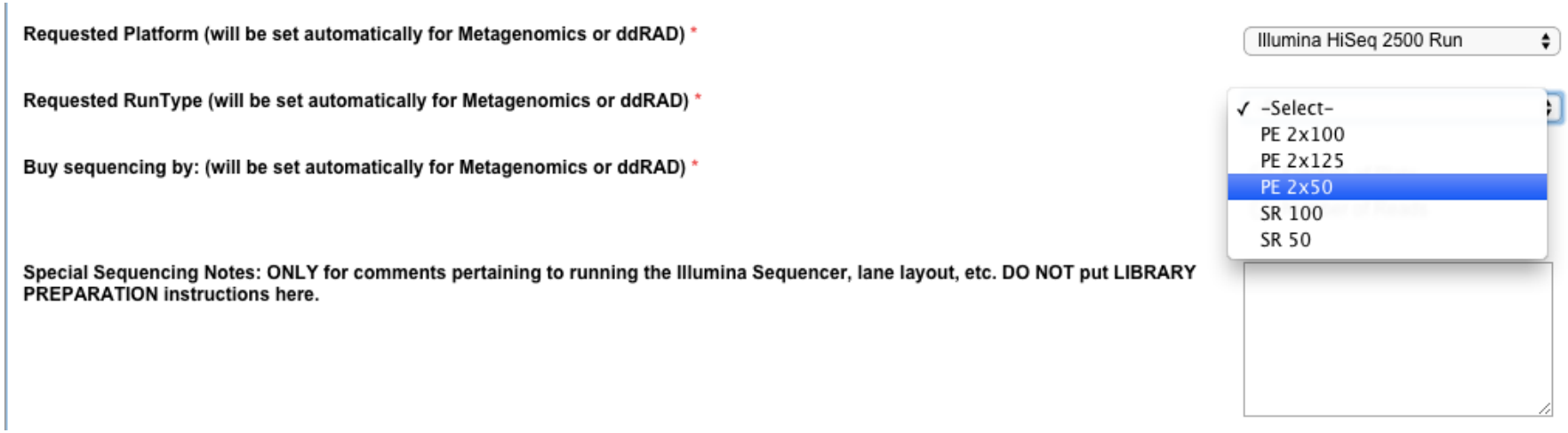
- All run types indicate (# of reads) x (read length)
  - e.g. “1x50” means single-end read, 50 bp long
- Any run type can be single- or dual-indexed

Sequencer run types & price



- All run types indicate (# of reads) x (read length)
  - e.g. “1x50” means single-end read, 50 bp long
- Any run type can be single- or dual-indexed

Sequencer run types & price



- All run types indicate (# of reads) x (read length)
  - e.g. “1x50” means single-end read, 50 bp long
- Any run type can be single- or dual-indexed

Sequencer layouts & capacity

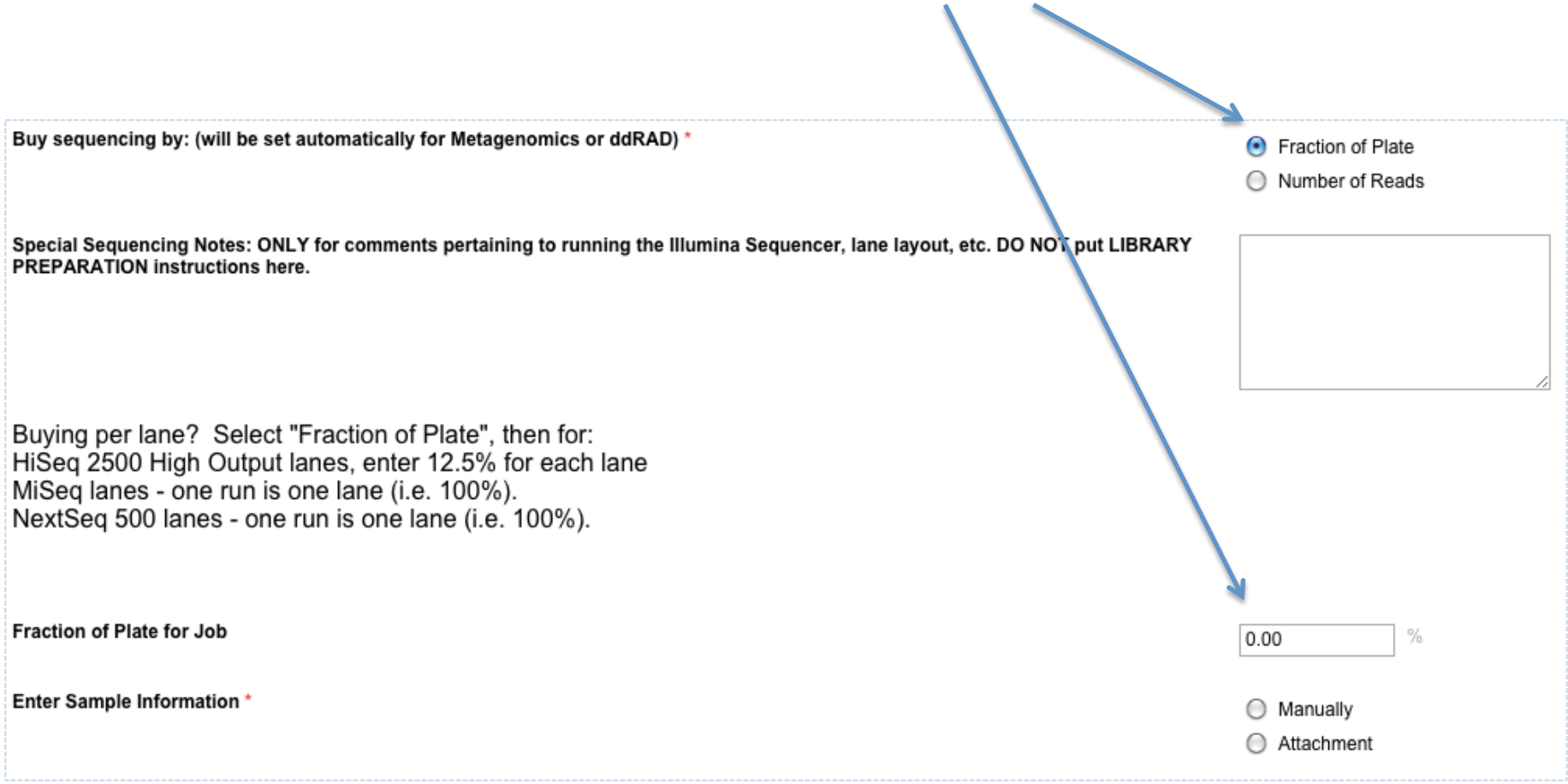
- Data capacity per run:
  - HiSeq: 8 lanes/run, 220 million reads per lane
  - NextSeq: 1 lane/run, 330 million reads per lane
  - MiSeq: 1 lane/run, 13 or 22 million reads per lane
- Barcode capacity per run:
  - Single-indexes: typically MAX of 12 per lane, unless prepared by the GSAF who have >48 barcodes available
  - Dual-indexes: typically >12, up to 400 per lane
  - Single- and dual-indexes can be run in the same lane IF they can be separated after the run

Purchasing “by read”

- The UT GSAF is known for selling “by the read” and not just “by the lane”
- What this means:
  - Target vs. Minimum:
    - We aim to sequence “Target”
    - If we fall below Minimum, we will automatically re-queue your sample for more sequencing
    - You will be billed for the LESSER of Target or what you actually receive
  - Minimum order per instrument:
    - HiSeq: 20 million read-pairs
    - MiSeq: 1 million read-pairs
    - NextSeq: > of 20 million reads per sample or 40 million reads per job

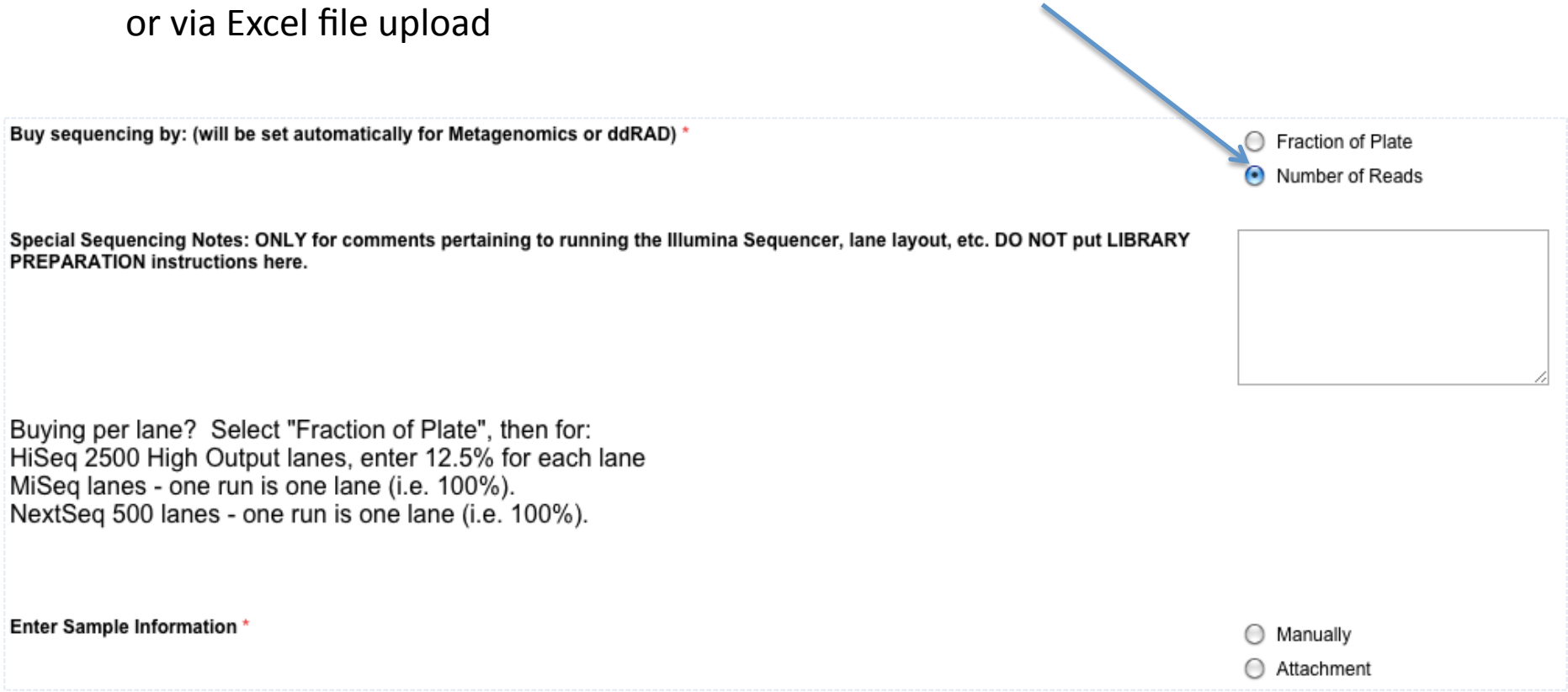
Purchasing “by read”

“Fraction of Plate” allows you to enter a % below



Purchasing “by read”

“Number of Reads” allows you to enter a number PER-SAMPLE on a later screen or via Excel file upload

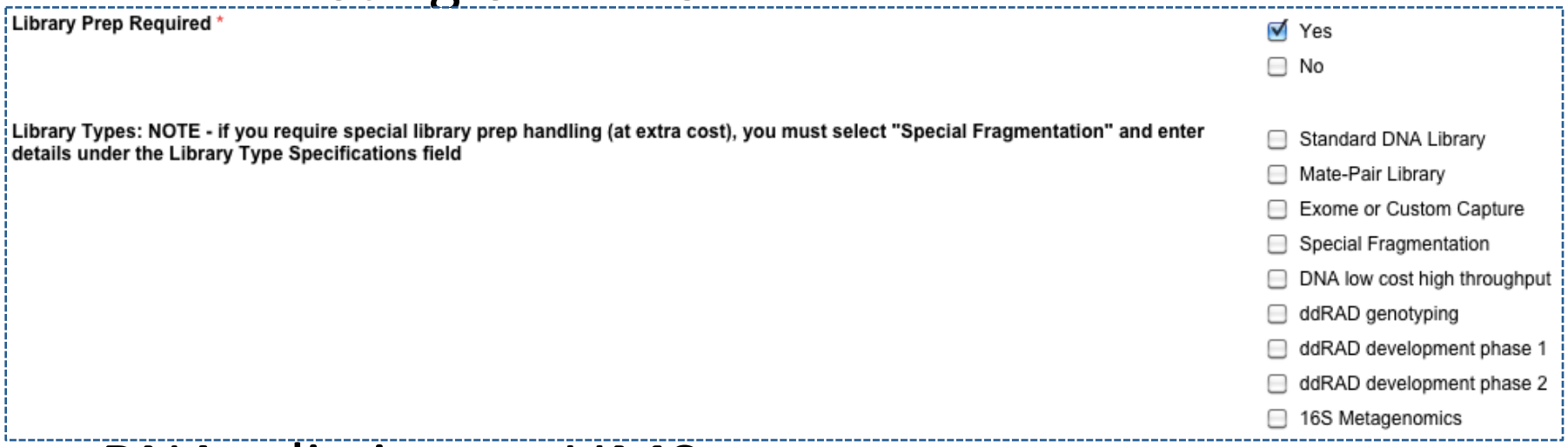


Library Types

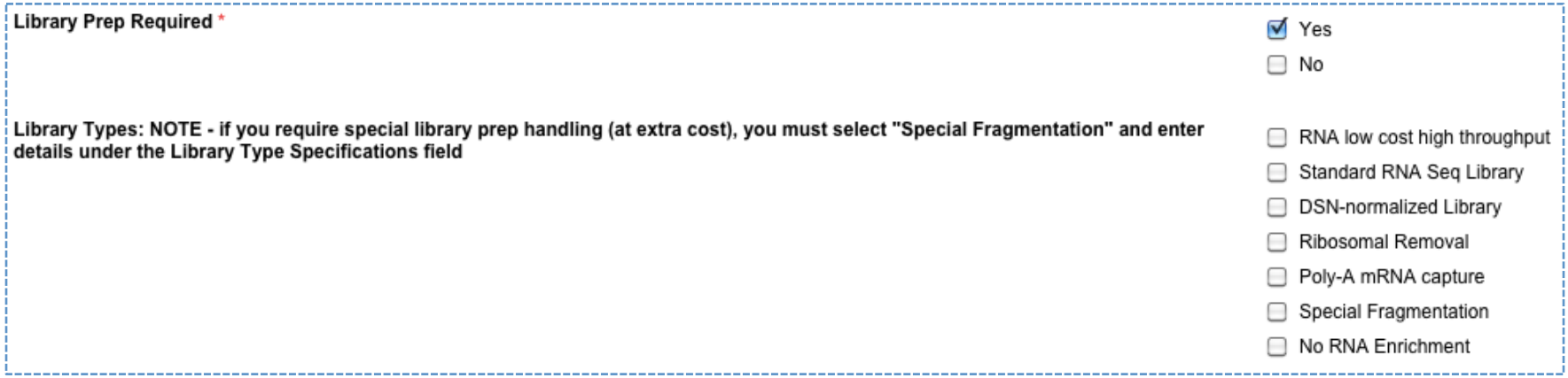
- DNA
  - Whole genome
  - Exome
  - ChIP
  - Amplicon
  - RAD
- RNA
  - Whole transcriptome – poly-A(+) or Ribosomal(-)
  - Small RNA
  - RIP

Library Types

- DNA – listing on LIMS

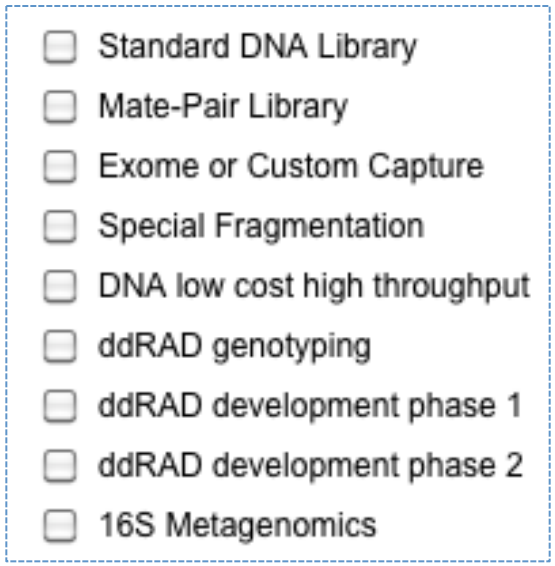


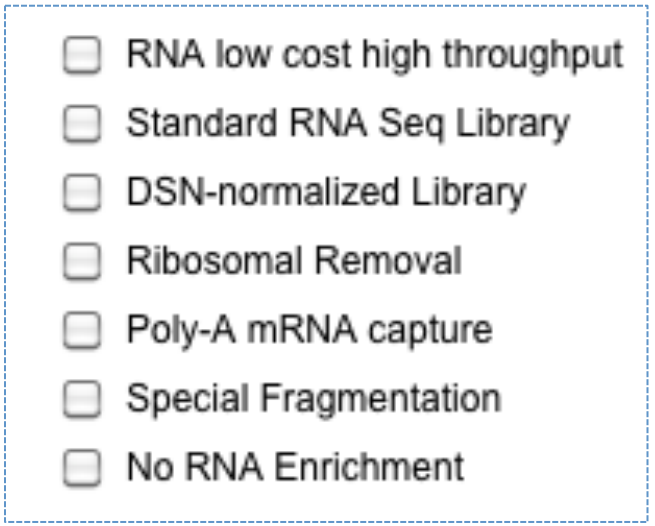
- RNA – listing on LIMS



Library Types

- DNA
  -
  -
  -
  -
  -
- RNA
  -
  -
  -



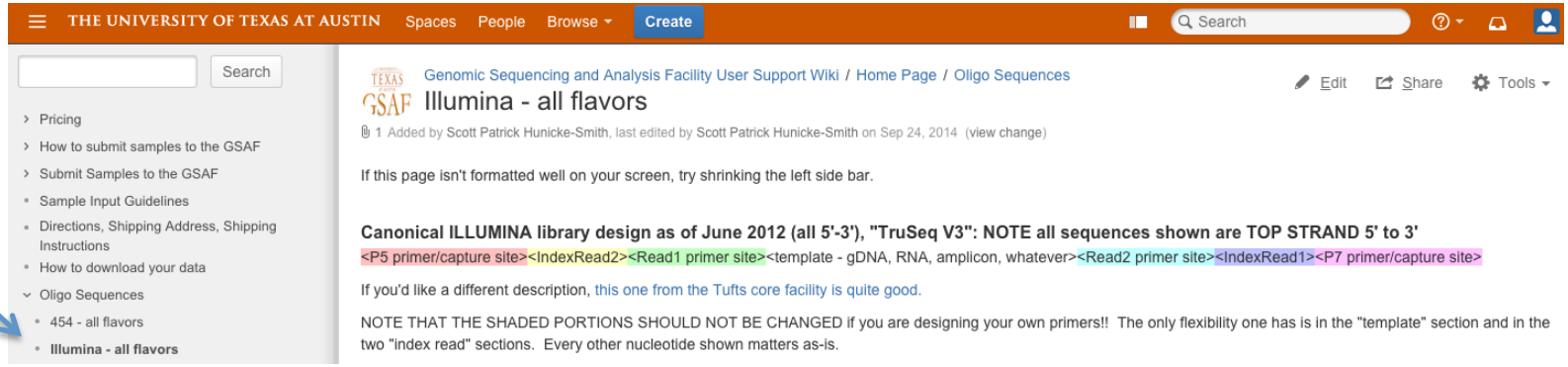


What Reagents do I use for Library Prep if I want to do it myself?

- There are **many options** to choose from when selecting library prep reagents
- The GSAF mainly uses New England BioLabs (NEB) for all their library prep reagents
  - Manuals are easy to follow
  - Reagents are well priced
  - You can purchase the reagents as a "kit" or as individual modules of the kit, which can be very helpful!
- We've also published our protocols in Wiley's Current Protocols:
  - PMID: 24984855 – DNA
  - PMID: 24733242 - RNA

Library Types & Sequencing

- Important concepts:
  - Single-end vs paired-end sequencing
  - Single-indexing, dual-indexing, and inline indexing



Genomic Sequencing and Analysis Facility User Support Wiki / Home Page / Oligo Sequences
Illumina - all flavors
Edit
Share
1 Added by Scott Patrick Hunicke-Smith, last edited by Scott Patrick Hunicke-Smith on Sep 24, 2014 (view change)
If this page isn't formatted well on your screen, try shrinking the left side bar.
Canonical ILLUMINA library design as of June 2012 (all 5'-3'), "TruSeq V3": NOTE all sequences shown are TOP STRAND 5' to 3'
<P5 primer/capture site><IndexRead2><Read1 primer site><template - gDNA, RNA, amplicon, whatever><Read2 primer site><IndexRead1><P7 primer/capture site>

- Platform descriptions and experimental design suggestions
- Dashboard
- Big Bio Job Board
- Companion Animal Interest Group: Information

to the dA insert tail so sequencing starts with the insert)

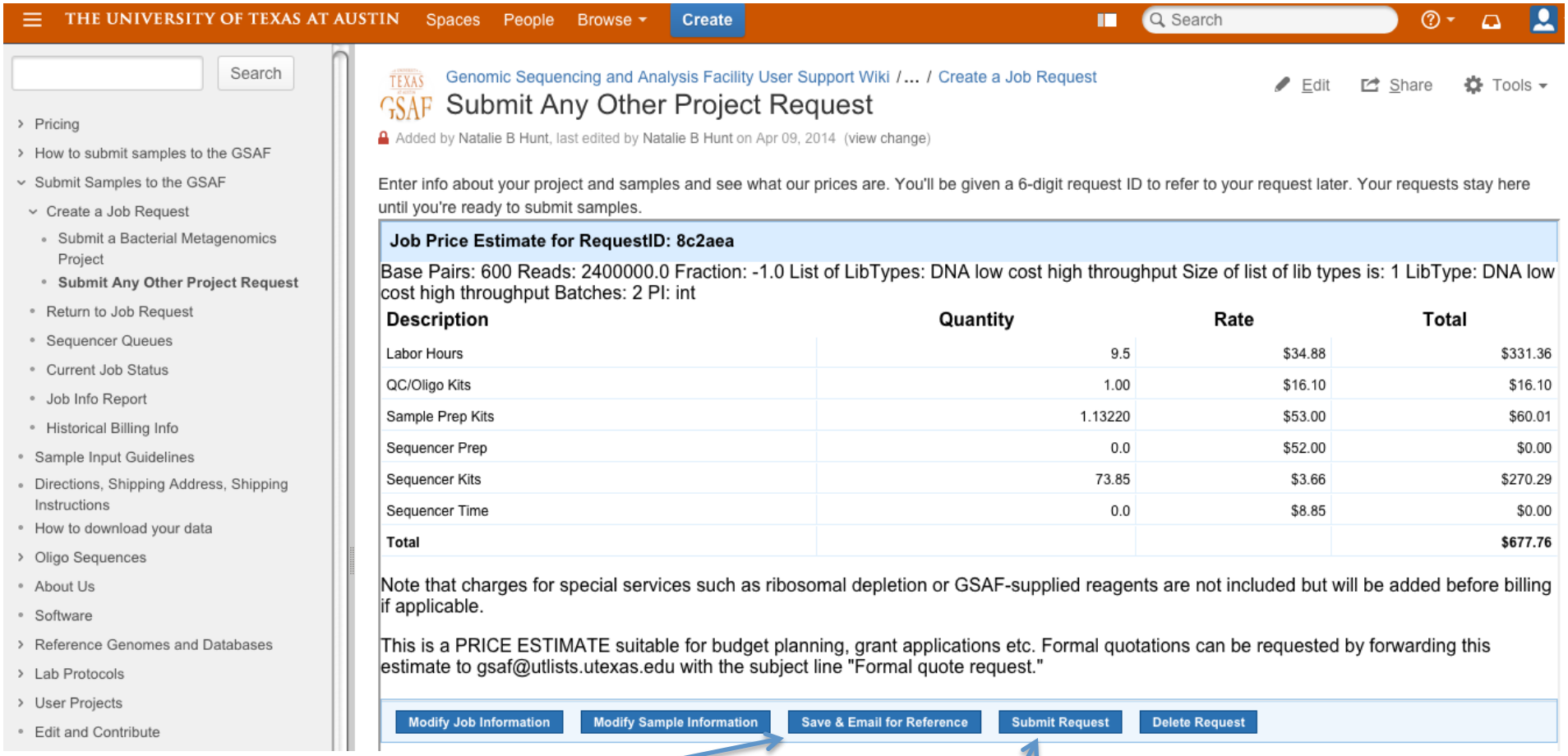
6. IndexRead1:

   The index sequence (usually 6 bp) - see many examples below in the **Barcodes** section. Within a lane, image analysis works best with as much base diversity as possible.
7. P7 PCR primer/flowcell capture site:

   ATCTCGTATGCCGTCTTCTGCTTG

How do I start a project with the GSAF?

- Job Submission example
  - Saving vs. Submitting your job



Save and Email – just for you!

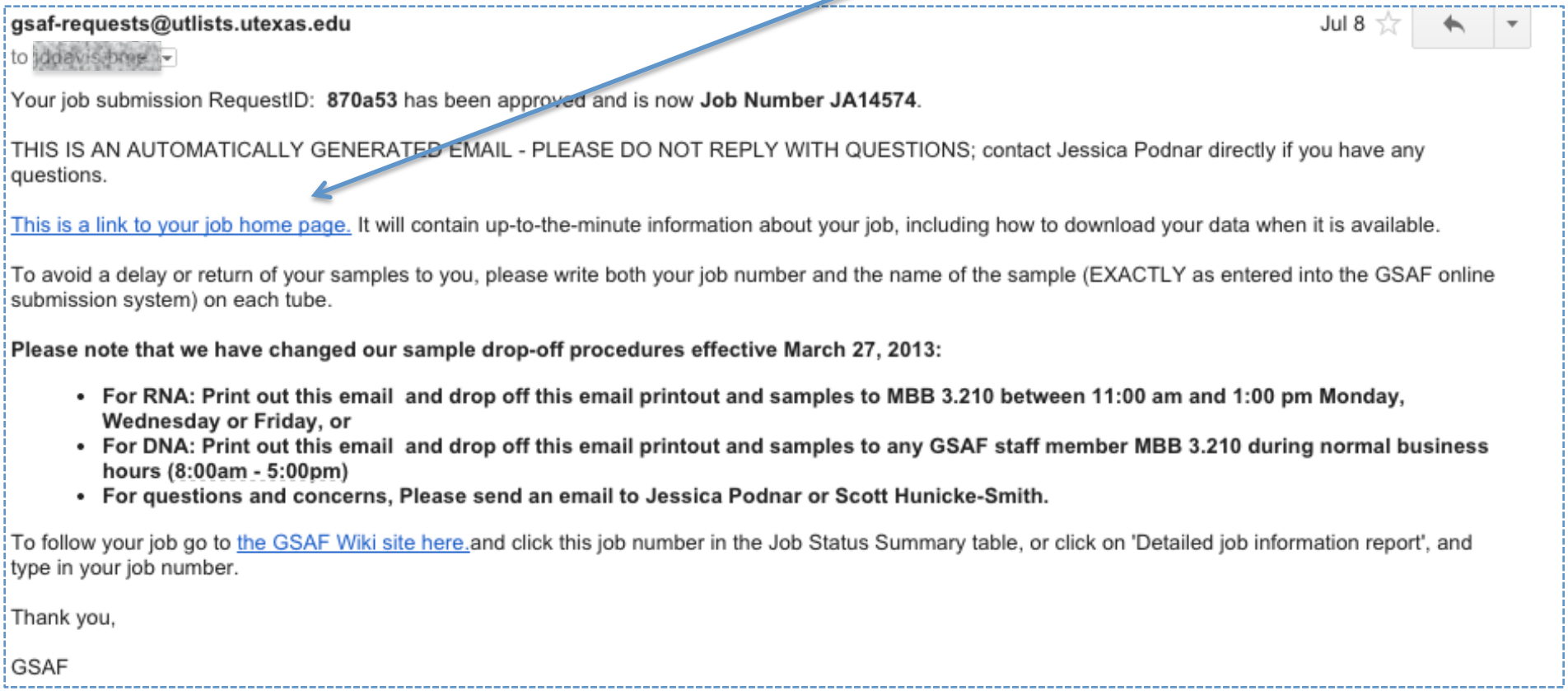
Submit Request – goes to GSAF for review/approval

How do I start a project with the GSAF?

- Job Submission example
  - Editing your job BEFORE submitting it
    - You cannot edit your job once you submit it

How can I tell where my project is?

- Your job submission email – link to your job home page



gsaf-requests@utlists.utexas.edu Jul 8

to jddavis@me

Your job submission RequestID: **870a53** has been approved and is now **Job Number JA14574**.

THIS IS AN AUTOMATICALLY GENERATED EMAIL - PLEASE DO NOT REPLY WITH QUESTIONS; contact Jessica Podnar directly if you have any questions.

This is a link to your job home page. It will contain up-to-the-minute information about your job, including how to download your data when it is available.

To avoid a delay or return of your samples to you, please write both your job number and the name of the sample (EXACTLY as entered into the GSAF online submission system) on each tube.

Please note that we have changed our sample drop-off procedures effective March 27, 2013:

- **For RNA: Print out this email and drop off this email printout and samples to MBB 3.210 between 11:00 am and 1:00 pm Monday, Wednesday or Friday, or**
- **For DNA: Print out this email and drop off this email printout and samples to any GSAF staff member MBB 3.210 during normal business hours (8:00am - 5:00pm)**
- **For questions and concerns, Please send an email to Jessica Podnar or Scott Hunicke-Smith.**

To follow your job go to the GSAF Wiki site here.and click this job number in the Job Status Summary table, or click on 'Detailed job information report', and type in your job number.

Thank you,

GSAF

How can I tell where my project is?

- Your job home page

GSAF Data System: Job Home Page for Job JA14227

This page contains up-to-the-minute information about your job. If a section is blank (e.g. QC Information, or Billing information), it means that we have not yet done that task for your job. This information is pulled direction from our production databases, so as soon as the information is available it will appear on this page.

Job ID: JA14227

Description: ddRAD protocol development

Sample receipt date: 17-Mar-14
Job completion date: 20-May-14

This project has 3 samples and there are 4 QC runs associated with this job.

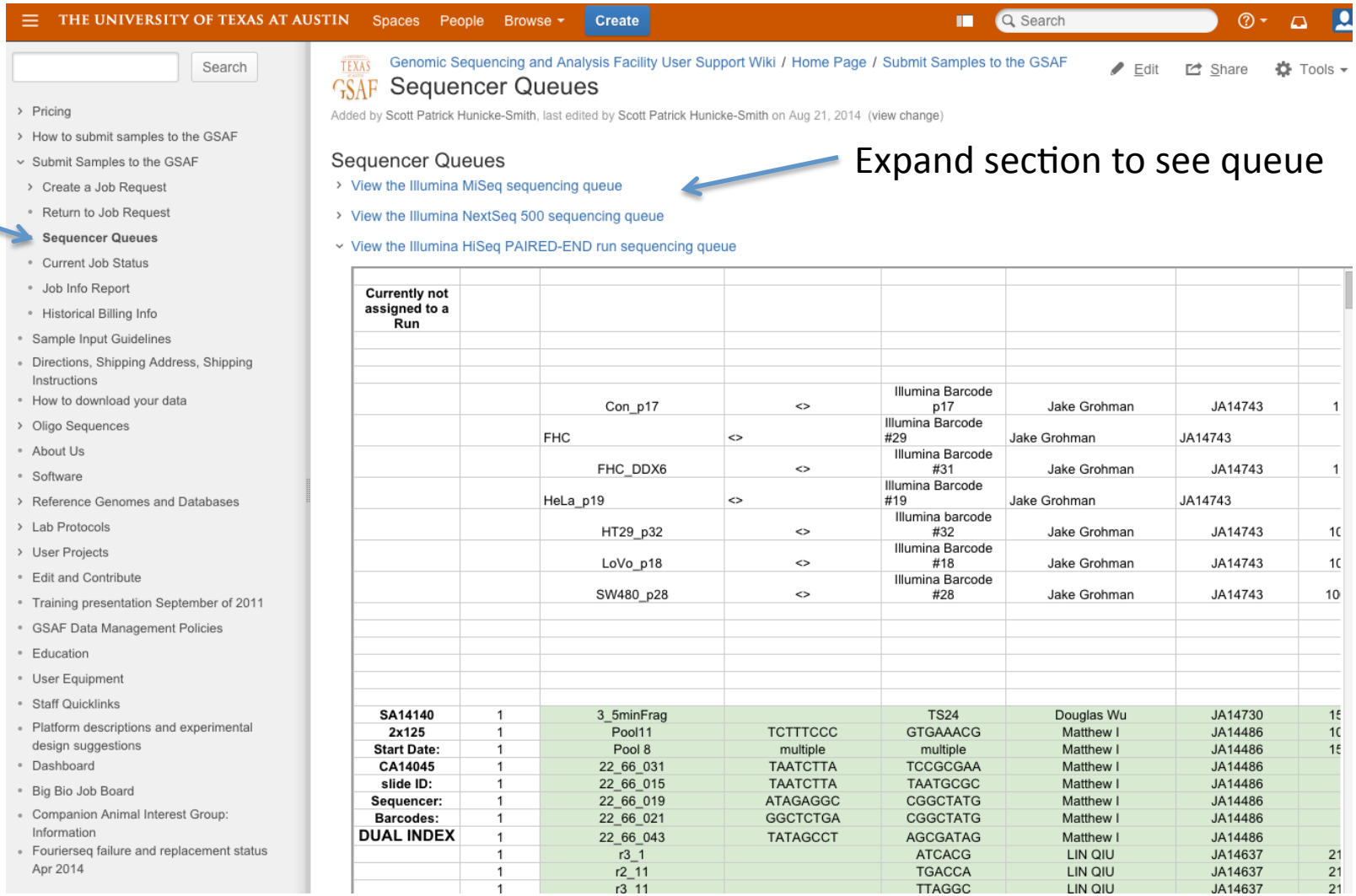
More information is available below - simply expand the sections to see more.

- ▶ QC Information
- ▶ Job Summary
- ▶ Billing Information
- ▶ Sequencing Data (expand to request new data access keys)

- ▶ Total sequencing data generated (expand for detail)

How can I tell where my project is?

- The sequencing queues



How do I get my data?

- Your data delivery email comes to you automatically

GSAF Data Key Available

New access keys have been generated for each of your data files. Anyone with access to your data keys will be able to download your data. We have limited the time window for download to 72 hours.

Access your data for JA14699 from sequencing run SA14130 here.

This link will be accessible for ten days from the date of this email.

You may request *new* access keys and find many details about your job on your job home page.

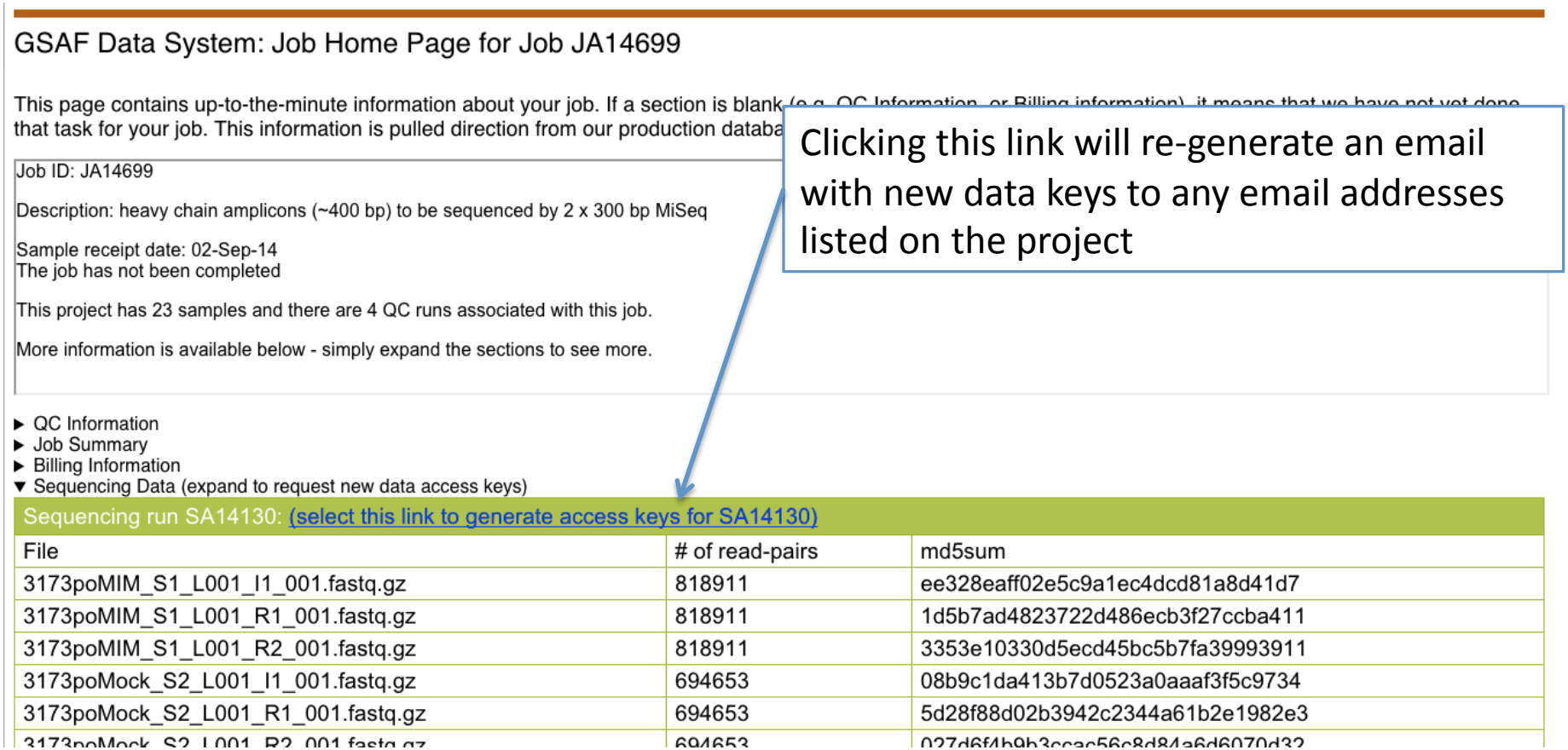
Please see this page for more information about downloading your data.

Sincerely,
GSAF

This email was auto-generated by data delivery system version 0.0.2 (alpha) of the UT GSAF.

How do I get my data?

- Your job home page – you can access ANYTIME



GSAF Data System: Job Home Page for Job JA14699

This page contains up-to-the-minute information about your job. If a section is blank (e.g. QC Information, or Billing information), it means that we have not yet done that task for your job. This information is pulled direction from our production databa

Job ID: JA14699

Description: heavy chain amplicons (~400 bp) to be sequenced by 2 x 300 bp MiSeq

Sample receipt date: 02-Sep-14
The job has not been completed

This project has 23 samples and there are 4 QC runs associated with this job.

More information is available below - simply expand the sections to see more.

Clicking this link will re-generate an email with new data keys to any email addresses listed on the project

▶ QC Information
▶ Job Summary
▶ Billing Information
▼ Sequencing Data (expand to request new data access keys)

Sequencing run SA14130: (select this link to generate access keys for SA14130)

File	# of read-pairs	md5sum
3173poMIM_S1_L001_I1_001.fastq.gz	818911	ee328eaff02e5c9a1ec4dcd81a8d41d7
3173poMIM_S1_L001_R1_001.fastq.gz	818911	1d5b7ad4823722d486ecb3f27ccba411
3173poMIM_S1_L001_R2_001.fastq.gz	818911	3353e10330d5ecd45bc5b7fa39993911
3173poMock_S2_L001_I1_001.fastq.gz	694653	08b9c1da413b7d0523a0aaaf3f5c9734
3173poMock_S2_L001_R1_001.fastq.gz	694653	5d28f88d02b3942c2344a61b2e1982e3
3173poMock_S2_L001_R2_001.fastq.gz	694653	027d6f4b9b3ccac56c8d84a6d6070d32

Frequently Asked Questions

Example Projects

- RNA-seq, poly-A mRNA enrichment for differential expression
- Bacterial whole-genome assembly
- ChIP-seq
- ddRAD

Thank you from the GSAF!

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