Library Prep and NGS Pricing Descriptions

Any given sequencing project is a combination of core services with UT-approved rates sized according to the details of the project. Prices shown below are estimates based on typical pricing. According to UT System and UT Austin policy, the GSAF bills for actual expenses which vary with the size and scope of each project.

To create and save a price estimate for future use, create a job request here. If you are seeking price estimates only, please do not submit the request until you have samples ready.

**Sequencing Pricing: NOTICE 12Jan18** Please Inquire about HiSeq 4000 lanes, we will be discontinuing these runs so space is limited and once all final lanes have been filled which should be by the end of February we will continue on with only HiSeq 2500 full lanes only, all partial sequencing jobs will be run on the NextSeq, this rule is being implemented to improve TAT’s.

Sequencing prices per lane (HiSeq) or per flow cell (NextSeq and MiSeq), Please email the GSAF if you are interested in a Rapid Run

<table>
<thead>
<tr>
<th>Platform</th>
<th>Run Type</th>
<th>Internal / UT</th>
<th>External Academic</th>
<th>External Commercial</th>
</tr>
</thead>
<tbody>
<tr>
<td>HiSeq 2500</td>
<td>SR 50</td>
<td>$1052</td>
<td>$1331</td>
<td>$1357</td>
</tr>
<tr>
<td>HiSeq 2500</td>
<td>SR 100</td>
<td>$1428</td>
<td>$1806</td>
<td>$1842</td>
</tr>
<tr>
<td>HiSeq 2500</td>
<td>PE 125</td>
<td>$2520</td>
<td>$3187</td>
<td>$3250</td>
</tr>
<tr>
<td>HiSeq 4000</td>
<td>SR 50 (1)</td>
<td>$1,043</td>
<td>$1,319</td>
<td>$1,345</td>
</tr>
<tr>
<td>HiSeq 4000</td>
<td>PE 150 (1)</td>
<td>$2,562</td>
<td>$3,239</td>
<td>$3,266</td>
</tr>
<tr>
<td>HiSeq 4000</td>
<td>SR 50</td>
<td>$1,697</td>
<td>$2,146</td>
<td>$2,173</td>
</tr>
<tr>
<td>HiSeq 4000</td>
<td>PE 75</td>
<td>$14,494</td>
<td>$18,327</td>
<td>$18,531</td>
</tr>
<tr>
<td>NextSeq 500</td>
<td>SR 75 H.O (1,3)</td>
<td>$2,302</td>
<td>$2,906</td>
<td>$3,060</td>
</tr>
<tr>
<td>NextSeq 500</td>
<td>PE 75 H.O. (1,3)</td>
<td>$3,826</td>
<td>$4,834</td>
<td>$4,988</td>
</tr>
<tr>
<td>NextSeq 500</td>
<td>PE 150 H.O.</td>
<td>$5,735</td>
<td>$7,250</td>
<td>$7,404</td>
</tr>
<tr>
<td>MiSeq</td>
<td>V2 - 300 cycles</td>
<td>$1,627</td>
<td>$2,054</td>
<td>$2,157</td>
</tr>
<tr>
<td>MiSeq</td>
<td>V2 - 500 cycles(1)</td>
<td>$1,771</td>
<td>$2,236</td>
<td>$2,339</td>
</tr>
<tr>
<td>MiSeq</td>
<td>V3 - 150 cycles</td>
<td>$1,463</td>
<td>$1,847</td>
<td>$1,950</td>
</tr>
<tr>
<td>MiSeq</td>
<td>V3 - 600 cycles(1)</td>
<td>$2,183</td>
<td>$2,758</td>
<td>$2,860</td>
</tr>
</tbody>
</table>

Notes:

1) These run types (bold) are available as shared runs on the GSAF "per-read" pricing schedule.
2) The HiSeq 4000 PE 2x75 is available only as full, 8-lane runs.
3) H.O. is Illumina's High Output run type. Mid Output is not available under standard pricing (inquire if interested).

Explanation of yields and the GSAF "per-read" pricing:

The UT GSAF is known for offering our "per-read" pricing option, so when applicable, clients are not required to purchase a full lane or a full flow cell of sequence data. This is a cost-effective way to do many NGS experiments, such as metagenomics, genotyping-by-sequencing, RNA-seq, and whole genome bacterial sequencing, among others.

However, "per-read" pricing is not available on every run type, and there are some constraints: **PLEASE NOTE CHANGES TO HISEQ**

- Minimum read counts per-sample are as follows:
  - Illumina MiSeq V2 or V3: 1 million reads
• Illumina NextSeq 500: the larger of 40 million reads per project or 20 million reads per sample.
• Illumina HiSeq 4000: 20 million reads per single sample; 10 million reads/sample if submitting 2 or more.

• All indexes (barcodes) must be pre-screened for potential conflicts. (See the Important Note below.)
• Samples with dual-indexes may not be run in partial lanes on any platform - they must be run in their own whole lane.

• We commonly achieve 210 to 230 million reads per lane on GSAF-prepared libraries (generally lower on SE runs) or more from HiSeq 2500 high output lanes and expect to achieve >240 million for HiSeq 4000 lanes.
• We do NOT guarantee a minimum yield of sequences from a lane with customer-prepared libraries. If you request a full lane of sequence, you will be charged for a full lane of sequence unless something has gone wrong with the instrument or entire flow cell (i.e. the problem is instrument related and not library related).

To calculate "per-read" pricing, we assume:
• a yield of 240 million reads per lane from each HiSeq 4000 lane
• a yield of 200 million reads per lane from each HiSeq 2500 lane
• a yield of 13 million reads per lane from MiSeq V2 and 22 million reads per lane from MiSeq V3
• a yield of 330 million reads per run from NextSeq 500, all run types.

When you request "per-read" pricing, you specify a minimum and a target read count, there is no guarantee but we always try our best to at least hit the minimum, if we do not hit the min you are only charged for the number of reads received. Your maximum charge (and thus the estimate you will receive when you submit your project) will be based on the target read count.

IMPORTANT NOTE:
The choice of indices matters. If you are constructing your own finished libraries, we recommend contacting the GS AF before assigning indices to your samples. If, for example, you use TruSeq Index 1 to 6 repeatedly, the wait time may be significantly longer since they will conflict internally and with other projects. Please consider using other indices or have the GS AF prepare your libraries.

Libraries with standard TruSeq or Nextera indices may be multiplexed in a lane. Libraries created by the GS AF can be multiplexed up to 60 samples per lane for DNA or RNA, 672 samples per lane for 16S-based metagenomics or ddRAD.

Libraries with in-line barcodes must be provided pooled by the customer and run in their own lane.

Payment and turn-around time

All finished libraries submitted for sequencing will bear a nominal charge for quality assessment and sample handling. This charge is per tube submitted. Lane prices shown above include this charge for 1 sample per lane.

The "Internal UT Rate" is charged when payment comes from an account internal to UT Austin. All other service is charged at the External Rate.

Typical turn-around times are 4-6 weeks for HiSeq 4000 runs, and 2-4 weeks for NextSeq 500 and MiSeq runs.

Library Preparation Pricing:

High throughput DNA and RNA library preps:

**DNA** low cost high throughput - wide size range, ~400 bp insert size. Minimum batch size of 8 samples per project. For fewer than 8 samples, your job will be queued until a minimum batch size of 8 is reached. Price includes indexing and final QC by BioAnalyzer and qPCR. See this web page for sample input guidelines.

**RNA** low cost high throughput - size range typically ~200 nt; "dUTP" directional protocol. Always processed in batches of at least 8; if you submit fewer than 8 samples, your job will be queued until a minimum batch size of 8 is reached. Price includes indexing and QC by BioAnalyzer on final lib only and qPCR. Please request a BA if you need to check the quality of the RNA before we proceed to library prep. See this web page for sample input guidelines.

Prices listed are per sample. For projects with 8 or less samples, a flat fee per project is charged as shown.

<table>
<thead>
<tr>
<th>Type of Library Preparation Procedure</th>
<th>Samples</th>
<th>Internal / UT Network</th>
<th>External Academic</th>
<th>External Commercial</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA High-Throughput with index</td>
<td>8 or more</td>
<td>$101</td>
<td>$127</td>
<td>$137</td>
</tr>
<tr>
<td>RNA directional WITH poly-A enrichment and index</td>
<td>8 or more</td>
<td>$206</td>
<td>$260</td>
<td>$283</td>
</tr>
</tbody>
</table>
RNA directional WITH ribosomal removal and index 8 or more $301 $381 $416
RNA directional with index - NO removal of ribosomal RNA 8 or more $141 $179 $206
Added fee/project for services with fewer than 8 samples 7 or less $331 $462 $551

Metagenomics assays on the Illumina MiSeq:

Metagenomics assay: This service starts with normalized (i.e. equal concentration and volume) gDNA and includes amplification using one of the GSAF provided primer sets chosen by the customer as shown on this web page. The service is calibrated to deliver at least 10,000 2x250 bp paired-end sequences from the Illumina MiSeq platform for at least 90% of the samples submitted, assuming there is reasonable amount of DNA to amplify. To minimize jackpot effects, PCR is performed in triplicate for each sample and re-pooled.

The library preparation is a two-step process, a gene-specific PCR then a second PCR to add the dual indexes. The PCR targets the bacterial 16S V4/V5 region, the 16S V4 region, or fungal ITS region, as specified by the client. Gel-based QC is performed on a sampling of 10% of the samples post amplification. Genomic DNA is expected as input; the GSAF does not provide DNA extraction services at this time. See this web page for sample input guidelines and note that DNA concentrations should be normalized before submission. The GSAF can normalize your sample concentrations for an additional charge.

Prices are listed per sample to include library preparation from genomic DNA and sequencing. For jobs with less than 176 samples, a flat fee per project is charged as shown. Prices vary slightly depending on the actual number of samples processed, so the prices are provided for reference and planning purposes. Submit a job request to get accurate, approved pricing.

<table>
<thead>
<tr>
<th>Number of Samples</th>
<th>Internal / UT Network</th>
<th>External Academic</th>
<th>External Commercial</th>
</tr>
</thead>
<tbody>
<tr>
<td>176 or more</td>
<td>$28.77</td>
<td>$36.25</td>
<td>$41.02</td>
</tr>
<tr>
<td>50-175 samples</td>
<td>Add $369 / project</td>
<td>Add $464 / project</td>
<td>Add $554 / project</td>
</tr>
<tr>
<td>&lt; 50 samples</td>
<td>Inquire</td>
<td>Inquire</td>
<td>Inquire</td>
</tr>
</tbody>
</table>

IMPORTANT CAUTION: The V4/V5 region assay we use is KNOWN to amplify Eukaryotic 18S rRNA genes as well as bacterial 16S V4/V5. If your sample may have Eukaryotic DNA (e.g. host), we have an alternative V4 region (only) assay but you MUST NOTE THIS in the Brief Description of your sequencing request.

Note: if you are submitting a mixed population (e.g. bacterial & fungal, or bacterial & eukaryotic host, etc.) you should assay your samples before submission to ensure you have the proper concentration of the target DNA, not simply total DNA. Samples submitted with insufficient target DNA may result in excessive primer-dimer or off-target products which produce no useful data. These will still be charged according to the pricing table.

GBS services

Genotyping by Sequencing (GBS) services

GBS genotyping: The library preparation service starts with pre-normalized (i.e. equal concentration and volume) DNA samples and includes double-digestion with one user-specified enzyme pair, user-specified size selection, adaptor ligation and purification. Sequencing targets at least 500,000 2x150 bp paired-end sequences from the Illumina HiSeq platform, with the expectation that 90% or more of the libraries will produce the targeted number of reads. This expectation is based on the customer providing the samples normalized and intact high MW DNA. This level of sequence depth assumes ~10,000 loci have been targeted; for more or less depth, please contact us for a custom quote. See this web page for sample input guidelines, noting that input DNA should be normalized before submission. The GSAF can normalize sample concentrations for an additional charge.

As described in this paper, by selecting one of several restriction enzyme pairs and a size range of the resulting fragments, one can "tune" the optimal coverage of a genome as suited to the scientific purpose.

The GSAF offers the following enzyme pairs for library prep: PhI-EcoRI, EcoRI-MspI, PhI-MluCI, NlaIII-MluC. Due to the sensitivity of the results to the digestion and size selection steps, we recommend both development and high-throughput processing be done by the same lab.

Expert Commentary from UT Prof. Dan Bolnick on this service...

The GSAF quotes a fixed price per sample, but this is a bit misleading. You really want to determine your depth of coverage and number of desired SNPs. Our calculation runs as follows:

- $X = \text{number of SNPs desired}$
- $\pi = \text{polymorphism frequency in the focal population(s) = per-site probability there is a SNP at that site}$
Y = number of bases of genome sequence needed to get X bases on average = X / π
FL = fragment length chosen (typically ~200 to 300 bases)
F = number of different fragments you need to sequence = Y / FL
D = desired mean depth of coverage for each (best to aim high: 15 to 30 reads per fragment gives you good success at identifying heterozygotes, if that matters to you)
Number of reads needed per individual = D * F
N = number of individuals

1. # reads for the whole project = N * D*F
2. Divide this last by the number of reads you get per illumina HiSeq lane (varies a bit) and you have a target number of lanes for sequencing
3. Best to err on the high side for each individual

Back to the fixed price that our GSAF quotes: in our experience, this is too much sequence data for QTL mapping and too little for association mapping. We are budgeting $10 per sample for our QTL needs aiming for at least 2000 SNPs, and $50 per sample for our association mapping projects that require >20,000 SNPs.

For the ddRAD library preparation and sequencing service the user specifies the enzyme pair, size range and sequencing platform. Pricing shown below aims for 500,000 reads per sample using a Paired End 150 on either a NextSeq of HiSeq 4000. Not all libraries will achieve the targeted 500,000 reads, if you need must meet a minimum of 500,000 reads per sample please request a higher target number of reads, i.e. 700,000 for target and 500,000 for the minimum. Otherwise please note that we will consider reads counts around 300K-500K acceptable if the DNA was high quality and normalized. **DNA that is not normalized, not pure or degraded will have very negative effects on the distribution of reads.**

<table>
<thead>
<tr>
<th>Service Description</th>
<th>Samples</th>
<th>Internal / UT Network</th>
<th>External Academic</th>
<th>External Commercial</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddRAD Prep and Sequencing; HiSeq 4000 * PE 2X150</td>
<td>176 or more</td>
<td>$29.05</td>
<td>$36.61</td>
<td>$41.02</td>
</tr>
<tr>
<td>ddRAD Prep and Sequencing; NextSeq 500 PE 2X150</td>
<td>176 or more</td>
<td>$32.41</td>
<td>$40.86</td>
<td>$45.45</td>
</tr>
<tr>
<td>ddRAD on either HiSeq or NextSeq platforms</td>
<td>50-175</td>
<td>Add $369 / project</td>
<td>Add $464 / project</td>
<td>Add $554 / project</td>
</tr>
</tbody>
</table>

Prices will vary slightly depending on the actual number of samples processed. For jobs with less than 176 samples, a flat fee per project is charged as shown. Submit a job request to get accurate, approved pricing.

**BioAnalyzer Quality Control**

Agilent BioAnalyzer traces on final libraries are included in sample preparation prices. All other Agilent BioAnalyzer assays are a flat $15.90 per sample for internal customers and $19.57 for external academic customers.