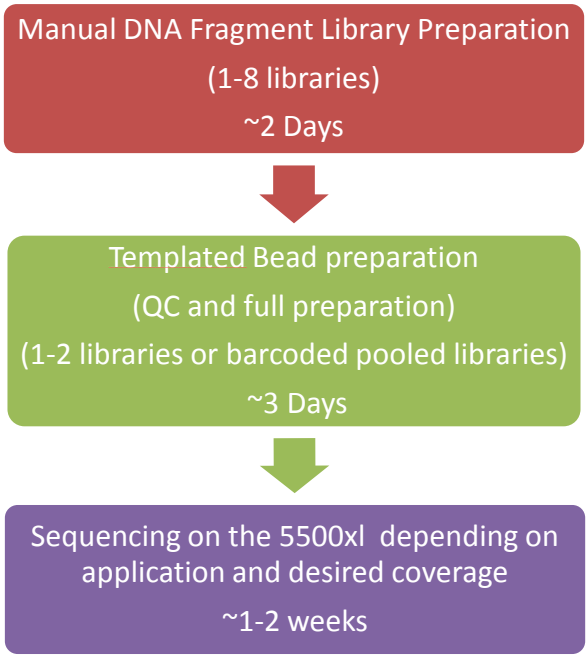
The information below is estimated for a SOLiD (5500xl) system in use at NIST. The discussion is to provide a 'snap shot' of the workflow and not intended to be a definitive cost/time analysis for NGS.

How much DNA do you need to prepare a library?

10 ng - 5 ug

How much time does it take to prepare a library and sequencing? Are robotics required?

To shear and prepare DNA with sequencing adaptors and barcodes (if desired) would take about 2 full days. You could comfortably prepare 8 different libraries in 2 days, 1- 8 libraries takes about the same amount of time. Following library preparation one can pool barcoded libraries to reduce cost per library. If this is done instead of running 1 library/ lane a pool of 12 libraries over 12 lanes could be run. Templated beads for sequencing would then need to be prepared. There are initial small scale bead preparations, two titration points, that are done to determine what concentration of library gives the best beads, that is the most monoclonal beads. Once these beads are made there is a short run on the sequencer to image the beads to look at their quality. Once this is done a large scale emulsion PCR is performed to prepare beads for a full sequencing run. The whole bead preparation process takes about 3 days for 1-2 libraries or pools. The bead enrichment step in this process is really a bottle neck. You can only do one sample/pool at a time and the complete process on that instrument takes about 6 hours. Below is a summary of the work flow and times for what we do here:



Robotics are not required but they do have methods for a few robots to automate portions of library preparation. They also have a new instrument ($50K), the Library Builder, which significantly reduces hands on time and can generate up to 26 libraries in 24 hours, so in all with quantitation about 1.5 days on this instrument. The bead preparation is done on proprietary instrumentation (~$45K). It is possible to do bead prep manually but laborious and variable.

What is the cost for reagents?

Below are the big ticket items. There are supplies needed for size selection, quantitation and various consumables that are not listed. These ancillary supplies probably total <$1000.

Kit	Cost per Kit	Units per Kit	Cost per Library
Core Library Reagent Kit	$1000	12	$83
Library Standard Adaptors	$450	12	$38
Library Barcode Adaptors	$1000	12	$83
		Library Prep Total per Library	***$121 - $166***
Titration Bead Preparations	$2600	4	$1300
Sequencing Reagents	$250	1-4	$63
Full Bead Preparations (varies depending on desired coverage)	$3800-$5800	4	$1000 - $1500
		Bead Prep Total per Library or Pool*	***$2363 - $2863***
Basic Forward Sequencing Reagents	$3300	12 lanes	$275
Forward Sequencing Primers	$160	12 lanes	$13
Forward Sequencing Primers for BC	$160	12 lanes	$13
Flow Chip and Deposition Supplies	$810	6 lanes	$135
Instrument Buffers	$700	12 lanes	$58
		Sequencing Total per Lane	***$494***

*You could significantly reduce your cost per library if libraries were pooled and pool was run over multiple lanes. It all depends on the application and coverage you need if this is practical.

**Cost of sequencing would vary depending on desired coverage. Increased coverage = increased number of lanes

NIST Multiplexed Biomolecular Science Group – 'Ask an NGS expert' feedback 01-31-2012

How much data is generated by a run (X GB)?

90 Gb (Giga bases) / run at capacity = 30X coverage of human genome (3 Gb)

How long does it take to transfer the data to a workstation?

It typically takes ~1 day to transfer data from a run between the instrument and the compute cluster for analysis (basically from one room to another), transferring to the cloud for analysis would probably take significantly longer. You can set up data transfer during the run to save time, we just haven't done this yet.

How long does it take to assemble the data?

Since a human reference genome is available, reads from the instrument are generally aligned to the reference genome, which generally take about 1 day on a cluster. Alignment is followed by post-processing and variant calling, which often take an additional 1 to 3 days for established algorithms. Though much less common, it is also possible to perform *de novo* assembly of parts of the human genome, which can take between 40 – 87 h depending on the amount of data, type of computational algorithm, and computational power available (Ruiqiang Li, Hongmei Zhu, Jue Ruan, et al. "De novo assembly of human genomes with massively parallel short read sequencing" Genome Res. 2010 20: 265- 272.) Some algorithms also combine an initial alignment step with local assembly around variants.

What infrastructure is needed to analyze the data (servers, IT/programming experts)?

It depends on the type of analysis, a lot of companies provide software for data analysis, you shouldn't need to be an IT/programming expert. Having a local computing cluster is helpful, but this would require a certain comfort level with UNIX and requires capital investment and IT support. An alternative is a cloud-based analysis option like Amazon's EC2 service.

Is the software for analysis 'custom made' or are there commercial packages for sequence analysis?

Both, it depends on the application and how much you trust the commercial packages. Every sequencing instrument manufacturer provides analysis software. If you want to compare data from different instruments then commercial third-party tools and open-source command-line tools are widely available, but results may vary.