Cost per Megabase/Genome

Considering the fact that the Human Genome Project cost around $2.7 billion dollars, the cost of sequencing is a major limiting factor when considering projects(2). Utilizing technologies that sequence at lower costs enable researchers to take on bigger projects and obtain more data3. Therefore, the cost it takes to obtain one megabase (1 million bases) of data is heavily considered by researchers when choosing what kind of sequencing technology to use.

Cost per megabase has been used for years as a standard price measurement for DNA sequencing technology. It is used by the National Human Genome Research Institute in their widely known and important benchmark graphs illustrating the decreasing costs associated with DNA sequencing4. This measurement captures all of the direct ‘production’ costs of producing the raw sequencing data. These production costs include (4):

1. Labor, administration, management, utilities, reagents, and consumables
2. Sequencing instruments and other large equipment (amortized over three years)
3. Informatics activities directly related to sequence production (e.g., laboratory information management systems and initial data processing)
4. Shotgun library construction (required for preparing DNA to be sequenced)
5. Submission of data to a public database
6. Indirect Costs as they relate to the above items

These are all important costs associated with producing DNA sequence data and should be captured in a cost per megabase measurement. However, many reports (often by the companies themselves) don't take all these costs into consideration and as a result price per final genome is sometimes a more accurate measure of sequencing costs. The cost per megabase Table 1 below provides information comparing cost per million bases for 454, Illumina, SOLiD, and 3730xl5.

Price per genome is an important measurement and is widely used. For instance, much hype has been made for the $1,000 dollar genome and this is the goal set by the National Human Genome Research Institute. This information is usually supplied in addition to the price per megabase and both will be used in our report as price comparators.

The price of purchasing the actual sequencing machine is a critically important factor for those institutions looking to be able to sequence in-house. However, for others looking to outsource their DNA sequencing needs, this cost is not as important as the cost per megabase and cost per genome. So, for the scope of this report we will refrain from incorporating the cost of the actual machines into our comparison.

Preliminary Analysis

Sources of MinIon data:

Since the Oxford nanopore technology is so new, there is much less data on its capabilities compared to the more established sequencers. The public’s main source of unbiased data comes from a handful of published studies done by “early access” customers. The early access program allowed a small number of selected labs, institutions and individuals to purchase the nanopore sequencer in early spring 2014. Besides giving a brief but useful explanation of the minION sequencer itself, these studies provide unbiased information on the data this machine produces and represent our main source of information regarding theprice, speed, and accuracy of this technology.

Results May Vary

The results of a sequencing run vary by a lot of factors. Because MinIon is such a new technology many of its protocols and reagents are in flux. This is especially true of the critical flow cells that form the basic consumable reagent of the MinIon. Early access users were given a generation of flow cells called R6. Within a short time improvements were made and versions R7 (release in July 2014) and R7.3 (release in September 2014) were made available. The advanced chemistry of these flow cells, while outside the scope of this report, plays a huge role in the quality of the results. This is demonstrated by certain studies which used older R6 concluded the technology is “next to useless” while others using the R7 technology are able to produce vastly improved results.

One major limitation of this study (that was quick to get pointed out by the creators of Oxford Nanopore) is that it uses an outdated flow cell version.

Official results (MinIon):

Internally - 1gigabase of data per run

Early users – 505 Megabases per run

Max read length – 50kb with comparable read quality throughout

“Chemistry” of nanopore is related to R7, R7.3, and R8 iterations.

Shipment is problem

R6 was given originally to early access users but was replaced by R7 now R7.3. R8 is a larger change

2,000$ for a flow cell (one run)/ 500$ for library prep (prepares several samples)

Error types – insertions and deletions (particularly isnertions that produce spurious data)