nature reviews genetics

https://doi.org/10.1038/s41576-023-00600-1

Review article

Beyond assembly: the increasing flexibility of single-molecule sequencing technology

In the format provided by the authors and unedited

Supplementary Note - Estimating the cost of sequencing

Flow cell cost

In parts of this review, we refer to the cost of sequencing on various platforms. Estimating the cost of an "average" sequencing run is often difficult due to many factors including variable levels of output and the rapid advancement of sequencing technologies. Additionally, costs are often dependent on agreements with the end users and how many flow cells are purchased at one time. Costs are also often advertised by manufacturers, however those numbers reflect maximum output conditions and ideal usage, which may not reflect the actual cost that an average user can expect to achieve.

Logsdon *et al*., estimated the cost of the latest Illumina, PacBio, and ONT sequencing devices[1]. However, costs have changed since then. So, here we attempted to estimate the cost of flow cells for the platforms we reference. We approached this from the perspective of an end user that would be paying a sequencing core to sequence their samples. This is the norm for Illumina and PacBio sequencing, while ONT sequencing can be more readily performed at the benchtop. However, in order to make these comparisons fair, we also estimate ONT prices based on the cost at a core.

To find price per flow cell, we did a non-exhaustive survey of the landscape of sequencing cores located in the United States (last accessed 02/15/23). To be as generous as possible, we focused on the prices of the highest output versions of each platform that are currently available. For Illumina, this is the NovaSeq 6000 S4 flow cell with 2x150 bp (or 300 cycle) sequencing. For PacBio, this is the Sequel II device run with 30 hour movies on SMRT Cells 8M. For ONT, this is sequencing performed on an ONT PromethION device with a PromethION flow cell. We also surveyed ONT MinION prices for completeness. Prices were used if they were publicly available (see Supplemental Table for links to prices). The number of sequencing cores that offer NovaSeq S4 sequencing far outnumbers the number of cores that offer PacBio or ONT sequencing, however we expect this gap to be decreased in the coming years. For the cost estimates, we used the "internal" prices listed by the cores.

In order to estimate output, we used the numbers provided in Logsdon *et al*.[1] (Table 1 in that publication). Again, to be as generous as possible, we use the highest value in the range of "mean" outputs for each platform.

The results of this non-exhaustive survey are in the Supplemental Tables and cost calculations are summarized below. For the purpose of this review, we will use the numbers below, however we expect these numbers to change drastically in the coming years.

Table 1. Flow cell cost and cost per Gb

Platform	Flow cell	Median cost at cores in US (USD)	Cores surveyed	Maximum average output (Gb)[1]	Cost per Gb (USD)
Illumina NovaSeq 6000	S4 (300 cycle)	$18,448.00	33	3000	$6
PacBio Sequel II	SMRTCell 8M	$1,942.00	13	30	$65
ONT MinION	MinION	$1,079.18	9	20	$54
ONT PromethION	PromethION	$1,700.00	9	100	$17

Estimating cost of concatemerization methods

Multiplexed Arrays Sequencing of isoforms with PacBio sequencing (MAS-ISO-seq)

MAS-ISO-seq is a concatemerization method that seeks to increase the throughput of PacBio Iso-seq[2]. The authors report obtaining between 35-40 million full-length non-chimeric (FLNC) transcripts per SMRT Cell 8M flow cell on a Sequel IIe instrument. While the authors report an increase in throughput, it is estimated from the number of circular consensus sequencing (CCS) reads in MAS-ISO-seq experiments. In order to get a better estimate of how this method compares to Iso-seq, we used Sequel II SMRT Cell 8M Iso-seq FLNC transcript counts taken from data generated in Logsdon, *et al*.[3]. These experienced PacBio users obtain between 1.55-2.19 million FLNC transcripts from a single flow cell. We believe these represent "real-world" numbers and will give the most accurate estimate of what to expect from Iso-seq.

Sampling Molecules Using Re-ligated Fragments (SMURF-seq)

SMURF-seq is a concatemerization method that seeks to increase the throughput of sequencing short fragments (~200 bp) with ONT sequencing[4]. The authors performed a non-concatemerized run of short reads on an ONT MinION sequencer, obtaining 2.58 million reads with their "diploid" sample (mean read length ~600 bp). They also performed two SMURF-seq runs on MinION flow cells with their "diploid" sample, recovering 7.28 and 7.55 million reads. While they also sequenced other samples, the "diploid" sample was the only sample where both non-concatemerized and concatemerized experiments were performed. Therefore, we decided to use the "diploid" numbers for our analysis, taking the average of the two SMURF-seq runs (7.4 million reads).

Table 2. Cost estimates for concatemerization methods

Application	Platform	Flow cell	Flow cell cost (USD)	Non-concatamer molecules (millions)	Concatemerized molecules (millions)	Non-concatamer cost (USD/million reads)	Concatamer cost (USD/million reads)
MAS-ISO-seq	PacBio Sequel II	SMRTCell 8M	$1,942.00	1.6-2.2	35-40	$883-$1,214	$49-$56
SMURF-seq	ONT minION	minION	$1,079.18	2.6	7.4	$415	$146

References

1. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **21**, 597–614 (2020).
2. Al'Khafaji, A. M. *et al.* High-throughput RNA isoform sequencing using programmable cDNA concatenation. Preprint at bioRxiv https://www.biorxiv.org/content/10.1101/2021.10.01.462818v1 (2021).
3. Logsdon, G. A. *et al.* The structure, function and evolution of a complete human chromosome 8. *Nature* **593**, 101–107 (2021).
4. Prabakar, R. K., Xu, L., Hicks, J. & Smith, A. D. SMURF-seq: efficient copy number profiling on long-read sequencers. *Genome Biol.* **20**, 134 (2019).