

### Comparison of PacBio HiFi and ONT costs for genome assembly

| Data types and coverage |   |            | Reagent cost                   |                               | Computational cost for data generation             |   |  |   |   | Computational cost for assembly and polishing |                                     |  |               |
|-------------------------|---|------------|--------------------------------|-------------------------------|--|---|--|---|---|---|-------------------------------------|--|---------------|
| Technology              | Coverage required for a "good" diploid human genome assembly <sup>a</sup> | Platform   | Reagent cost per fold-coverage | Total reagent cost per genome | CPU hours required for CCS per genome <sup>b</sup> | Cost per hour per CPU node <sup>c</sup> | GPU hours required for basecalling per genome <sup>d</sup> | Cost per hour per GPU node <sup>e</sup> | Total computational cost for data generation per genome | CPU hours required for assembly               | CPU hours required for polishing    | Total computational cost for assembly/polishing per genome   | Overall cost  |
| PacBio HiFi             | 30X (96 Gb)<br>[QV 50+]   | Sequel II  | \$138-275                      | \$3450-8250                   | 6000-8000  | \$1.008                                 | --   | --                                      | \$372-496   | ~5,000<br>(HiCanu) <sup>f</sup>               | 0                                   | \$313  | \$4135-9059   |
| Non-UL ONT              | 60X (192 Gb)<br>[QV 30-50]  | GridION    | \$160-1600                     | \$9600-96000                  | --   | --                                      | 90-100   | \$3.006                                 | \$271-301   | ~700<br>(Shasta) <sup>g</sup>                 | MarginPolish/HELEN;<br>Racon/Medaka | \$70 for assembly <sup>h</sup> ;<br>\$107 for MarginPolish/HELEN <sup>i</sup> ;<br>\$621 for Racon/Medaka <sup>g</sup> | \$10669-97099 |
|                         |   | PromethION | \$67-134                       | \$4020-8040                   |  |   |  |   |   |   |                                     |  | \$5089-9139   |

Non-UL ONT: Non-ultra-long ONT data; typically refers to data generated with standard ONT kits and protocols, resulting in a read length N50 20-50 kbp long

CCS: Circular consensus sequencing

<sup>a</sup> We define a "good" diploid human genome assembly as having a QV score  $\geq 30$  (or a base accuracy  $>99.9\%$ ) and a contig N50  $\geq 10$  Mbp

<sup>b</sup> Assumes 2000 CPU hours per SMRT Cell 8M

<sup>c</sup> AWS estimate: r5.4xlarge; 16 vCPUs; reported in Nurk et al., bioRxiv, 2020

<sup>d</sup> ONT basecalling is usually done live, where it would have \$0 additional cost. However, if not done live, high-accuracy basecalling would require the indicated GPU hours

<sup>e</sup> AWS estimate: p3.2xlarge; 1 Tesla V100; reported in Shafin et al., Nat Biotechnol, 2020

<sup>f</sup> Reported in Nurk et al., bioRxiv, 2020

<sup>g</sup> Reported in Shafin et al., Nat Biotechnol, 2020