genes

MDPI

Editorial

The Versatility of SMRT Sequencing

Matthew S. Hestand [1,2,*] and Adam Ameur [3,4,*]

1 Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45202, USA
2 Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45202, USA
3 Department of Immunology, Genetics and Pathology, Uppsala University, Science for Life Laboratory, 75025 Uppsala, Sweden
4 Department of Epidemiology and Preventive Medicine, Monash University, Melbourne 32901, Australia
* Correspondence: matthew.hestand@cchmc.org (M.S.H.); adam.ameur@igp.uu.se (A.A.)

Received: 20 December 2018; Accepted: 3 January 2019; Published: 4 January 2019

The adoption of single molecule real-time (SMRT) sequencing [1] is becoming widespread, not only in basic science, but also in more applied areas such as agricultural, environmental, and medical research. SMRT sequencing offers important advantages over current short-read DNA sequencing technologies, including exceptionally long read lengths (20 kb or more), unparalleled consensus accuracy, and the ability to sequence native, non-amplified, DNA molecules. These sequencing characteristics enable creation of highly accurate de novo genome assemblies, characterization of complex structural variation, direct characterization of nucleotide base modifications, full-length RNA isoform sequencing, phasing of genetic variants, low frequency mutation detection, and clonal evolution determination [2,3]. This Special Issue of Genes is a collection of articles showcasing the latest developments and the breadth of applications enabled by SMRT sequencing technology.

In basic science, SMRT sequencing enables studies into the molecular mechanisms of living cells at a new level of resolution. Perhaps the most advantageous feature of SMRT sequencing is that it facilitates sequencing of long DNA molecules at a very high accuracy. This has enabled the construction of high-quality reference genomes for a wide range of species, including new human genome assemblies, as presented in this special issue [4]. In addition, when SMRT sequencing is performed on native non-amplified DNA molecules, it is possible to access several layers of additional information hidden in the kinetic signals emitted by the polymerases during the sequencing reaction [1]. This kinetic information has been used to detect epigenetic modifications at base pair resolution and even phasing of methylation signatures in diploid organisms, as presented in this special issue [5]. Several important discoveries have already been made from this kinetic information, such as the widespread presence of 6mA modifications in the human genome [6], a modification that was previously thought to only be present in bacterial genomes. In addition to base modifications, SMRT sequencing data also enables us to study other events, such as DNA conformations [7]. Another aspect of SMRT sequencing is that it can be used to study RNA, and it is currently the only technology that can generate high-quality continuous sequences for full-length transcripts up to 10 kb or more. This makes it possible to study splicing variation at a completely new level of resolution [8,9]. SMRT sequencing is also paving the way for a new generation of computational approaches to explore and interpret these rich datasets [10–12]. In summary, SMRT sequencing is enhancing and even opening up new areas of basic research that were not accessible with previous sequencing technologies.

In terms of more applied areas, agriculture is benefiting from the advent of SMRT sequencing for examining important microbes, plants, and animals. SMRT sequencing, often with complementary technologies, has produced new genome assemblies for important crops, such as apples, maize, wine grapes, coffee, rice, black raspberries, asparagus, and cotton [11,13–20]. SMRT transcriptome sequencing has also given new insights into gene structures for rice, wheat, maize, sorghum, barley, and cotton [18,21–25]. Besides providing new references, these projects will improve plant cultivation,

such as identifying drought and disease resistant genes. Strategies to detect genetically modified organisms (GMOs) have also been proposed and enhanced with SMRT sequencing [26]. Animal genome assemblies have been produced for several agriculturally valuable species, such as the horse, cow, goat, chicken (including its transcriptome), and commercially important fish like haddock and cod [27–33]. These will lead to improvements in animal breeding, management, and disease resistance. Finally, sequencing of pathogenic bacteria and fungi affecting agriculturally important species is providing insight into the diversity and virulence factors of these pathogens, which in turn will assist in disease risk and management [34–36].

In environmental research, systematic efforts are ongoing to generate reference sequences for thousands of bacterial strains and microorganisms. Recently, this has expanded to the genomes of larger organisms, including vertebrates [37]. SMRT sequencing can also play an important role in ecology research, such as monitoring the composition of fungi in environmental soil or water samples [38,39]. New high-quality references for animal genomes, such as the great apes [40], will provide an invaluable resource for future evolutionary studies. During the last few years, new genome assemblies have also been created for several endangered species, including Hawaii's last crow species [41], aiding in conservation efforts.

Though SMRT sequencing has primarily been applied to basic research, there is a growing implementation for clinical utility [3,42]. The long and highly accurate reads produced from SMRT sequencing have proven to be useful to resolve complex and repetitive regions of the human genome associated with disease. SMRT sequencing is also a sensitive method to detect minor variants in cancer and infectious disease. Although most current methods are based on targeted sequencing, the value of long reads is also becoming apparent for whole-genome sequencing, which allows clinical professionals to resolve repeat expansions, transposable element insertions, and other complex genomic rearrangements that are difficult or even impossible to assess using short-read sequence data.

As we look forward, this technology will provide even longer and more accurate reads at a higher throughput. This will enable routine de novo assembly of both alleles in large diploid genomes, accompanied with tissue specific epigenetic DNA modification information. As a consequence, there will be a demand for a new generation of computational tools to compare complete genomes to each other, as opposed to a reference standard, and to phase genetic variants and epigenetic modifications over large chromosomal regions. By sequencing thousands of individuals with long reads, it will be possible to obtain a detailed picture of complex structural variation within large population cohorts of humans, as well as for other species. Such endeavors will give new insights to the function of the repetitive parts of the genome, and likely explain the cause of many genomic diseases. Looking further on the horizon, SMRT sequencing can be envisioned in combination with other technical advances, such as single cell sequencing to provide information on the epigenetic modifications occurring in single cells. SMRT sequencing has been steadily evolving since the commercial introduction of the technology in 2011. Just as short-read technologies have replaced microarrays and Sanger sequencing for a host of applications, we envision long-read single-molecule sequencing to replace short-read platforms for a majority of applications, as well as continue to evolve into new applications, throughout many different areas in the coming decade.

References

1. Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-time DNA sequencing from single polymerase molecules. *Science* **2009**, *323*, 133–138. [CrossRef] [PubMed]
2. Rhoads, A.; Au, K.F. PacBio sequencing and its applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [CrossRef] [PubMed]
3. Ardui, S.; Ameur, A.; Vermeesch, J.R.; Hestand, M.S. Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Res.* **2018**, *46*, 2159–2168. [CrossRef] [PubMed]

4. Ameur, A.; Che, H.; Martin, M.; Bunikis, I.; Dahlberg, J.; Hoijer, I.; Haggqvist, S.; Vezzi, F.; Nordlund, J.; Olason, P.; et al. De novo assembly of two Swedish genomes reveals missing segments from the human GRCh38 reference and improves variant calling of population-scale sequencing data. *Genes* **2018**, *9*, 486. [CrossRef] [PubMed]
5. Suzuki, Y.; Wang, Y.; Au, K.F.; Morishita, S. A statistical method for observing personal diploid methylomes and transcriptomes with single-molecule real-time sequencing. *Genes* **2018**, *9*, 460. [CrossRef]
6. Xiao, C.L.; Zhu, S.; He, M.; Chen, D.; Zhang, Q.; Chen, Y.; Yu, G.; Liu, J.; Xie, S.-Q.; Luo, F.; et al. N^6-methyladenine DNA modification in the human genome. *Mol. Cell* **2018**. [CrossRef]
7. Guiblet, W.M.; Cremona, M.A.; Cechova, M.; Harris, R.S.; Kejnovska, I.; Kejnovsky, E.; Eckert, K.; Chiaromonte, F.; Makova, K.D. Long-read sequencing technology indicates genome-wide effects of non-B DNA on polymerization speed and error rate. *Genome Res.* **2018**, *28*, 1767–1778. [CrossRef]
8. Tilgner, H.; Grubert, F.; Sharon, D.; Snyder, M.P. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9869–9874. [CrossRef]
9. Kuo, R.I.; Tseng, E.; Eory, L.; Paton, I.R.; Archibald, A.L.; Burt, D.W. Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. *BMC Genom.* **2017**, *18*, 323. [CrossRef]
10. Chin, C.S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E.; et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **2013**, *10*, 563–569. [CrossRef]
11. Chin, C.S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A.; et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [CrossRef] [PubMed]
12. Sedlazeck, F.J.; Rescheneder, P.; Smolka, M.; Fang, H.; Nattestad, M.; von Haeseler, A.; Schatz, M.C. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **2018**, *15*, 461–468. [CrossRef] [PubMed]
13. Wang, M.; Tu, L.; Yuan, D.; Zhu, D.; Shen, C.; Li, J.; Liu, F.; Pei, L.; Wang, P.; Zhao, G.; et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* **2018**. [CrossRef]
14. Daccord, N.; Celton, J.M.; Linsmith, G.; Becker, C.; Choisne, N.; Schijlen, E.; van de Geest, H.; Bianco, L.; Micheletti, D.; Velasco, R.; et al. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* **2017**, *49*, 1099–1106. [CrossRef] [PubMed]
15. Jiao, Y.; Peluso, P.; Shi, J.; Liang, T.; Stitzer, M.C.; Wang, B.; Campbell, M.S.; Stein, J.C.; Wei, X.; Chin, C.S.; et al. Improved maize reference genome with single-molecule technologies. *Nature* **2017**, *546*, 524–527. [CrossRef] [PubMed]
16. Minio, A.; Lin, J.; Gaut, B.S.; Cantu, D. How single molecule real-time sequencing and haplotype phasing have enabled reference-grade diploid genome assembly of wine grapes. *Front. Plant. Sci.* **2017**, *8*, 826. [CrossRef]
17. Tran, H.T.M.; Ramaraj, T.; Furtado, A.; Lee, L.S.; Henry, R.J. Use of a draft genome of coffee (*Coffea arabica*) to identify SNPs associated with caffeine content. *Plant. Biotechnol. J.* **2018**, *16*, 1756–1766. [CrossRef]
18. Stein, J.C.; Yu, Y.; Copetti, D.; Zwickl, D.J.; Zhang, L.; Zhang, C.; Chougule, K.; Gao, D.; Iwata, A.; Goicoechea, J.L.; et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat. Genet.* **2018**, *50*, 285–296. [CrossRef]
19. VanBuren, R.; Wai, C.M.; Colle, M.; Wang, J.; Sullivan, S.; Bushakra, J.M.; Liachko, I.; Vining, K.J.; Dossett, M.; Finn, C.E.; et al. A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. *Gigascience* **2018**, *7*. [CrossRef]
20. Harkess, A.; Zhou, J.; Xu, C.; Bowers, J.E.; Van der Hulst, R.; Ayyampalayam, S.; Mercati, F.; Riccardi, P.; McKain, M.R.; Kakrana, A.; et al. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* **2017**, *8*, 1279. [CrossRef]
21. Wang, M.; Wang, P.; Liang, F.; Ye, Z.; Li, J.; Shen, C.; Pei, L.; Wang, F.; Hu, J.; Tu, L.; et al. A global survey of alternative splicing in allopolyploid cotton: Landscape, complexity and regulation. *New Phytol.* **2018**, *217*, 163–178. [CrossRef] [PubMed]
22. Ren, P.; Meng, Y.; Li, B.; Ma, X.; Si, E.; Lai, Y.; Wang, J.; Yao, L.; Yang, K.; Shang, X.; et al. Molecular mechanisms of acclimatization to phosphorus starvation and recovery underlying full-length transcriptome profiling in barley (*Hordeum vulgare* L.). *Front. Plant. Sci.* **2018**, *9*, 500. [CrossRef] [PubMed]

23. Wang, B.; Regulski, M.; Tseng, E.; Olson, A.; Goodwin, S.; McCombie, W.R.; Ware, D. A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing. *Genome Res.* **2018**, *28*, 921–932. [CrossRef] [PubMed]
24. Clavijo, B.J.; Venturini, L.; Schudoma, C.; Accinelli, G.G.; Kaithakottil, G.; Wright, J.; Borrill, P.; Kettleborough, G.; Heavens, D.; Chapman, H.; et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* **2017**, *27*, 885–896. [CrossRef] [PubMed]
25. Dong, L.; Liu, H.; Zhang, J.; Yang, S.; Kong, G.; Chu, J.S.; Chen, N.; Wang, D. Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genom.* **2015**, *16*, 1039. [CrossRef] [PubMed]
26. Fraiture, M.A.; Herman, P.; Papazova, N.; De Loose, M.; Deforce, D.; Ruttink, T.; Roosens, N.H. An integrated strategy combining DNA walking and NGS to detect GMOs. *Food Chem.* **2017**, *232*, 351–358. [CrossRef]
27. Torresen, O.K.; Star, B.; Jentoft, S.; Reinar, W.B.; Grove, H.; Miller, J.R.; Walenz, B.P.; Knight, J.; Ekholm, J.M.; Peluso, P.; et al. An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genom.* **2017**, *18*, 95. [CrossRef]
28. Kalbfleisch, T.S.; Rice, E.S.; DePriest, M.S., Jr.; Walenz, B.P.; Hestand, M.S.; Vermeesch, J.R.; O'Connell, B.L.; Fiddes, I.T.; Vershinina, A.O.; Saremi, N.F.; et al. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun. Biol.* **2018**, *1*, 197. [CrossRef]
29. Koren, S.; Rhie, A.; Walenz, B.P.; Dilthey, A.T.; Bickhart, D.M.; Kingan, S.B.; Hiendleder, S.; Williams, J.L.; Smith, T.P.L.; Phillippy, A.M. *De novo* assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol.* **2018**. [CrossRef]
30. Bickhart, D.M.; Rosen, B.D.; Koren, S.; Sayre, B.L.; Hastie, A.R.; Chan, S.; Lee, J.; Lam, E.T.; Liachko, I.; Sullivan, S.T.; et al. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat. Genet.* **2017**, *49*, 643–650. [CrossRef]
31. Warren, W.C.; Hillier, L.W.; Tomlinson, C.; Minx, P.; Kremitzki, M.; Graves, T.; Markovic, C.; Bouk, N.; Pruitt, K.D.; Thibaud-Nissen, F.; et al. A new chicken genome assembly provides insight into avian genome structure. *G3* **2017**, *7*, 109–117. [CrossRef] [PubMed]
32. Torresen, O.K.; Brieuc, M.S.O.; Solbakken, M.H.; Sorhus, E.; Nederbragt, A.J.; Jakobsen, K.S.; Meier, S.; Edvardsen, R.B.; Jentoft, S. Genomic architecture of haddock (*Melanogrammus aeglefinus*) shows expansions of innate immune genes and short tandem repeats. *BMC Genom.* **2018**, *19*, 240. [CrossRef]
33. Thomas, S.; Underwood, J.G.; Tseng, E.; Holloway, A.K. Bench to basinet CvDC informatics subcommittee. Long-read sequencing of chicken transcripts and identification of new transcript isoforms. *PLoS ONE* **2014**, *9*, e94650. [CrossRef] [PubMed]
34. Dickey, A.M.; Loy, J.D.; Bono, J.L.; Smith, T.P.; Apley, M.D.; Lubbers, B.V.; DeDonder, K.D.; Capik, S.F.; Larson, R.L.; White, B.J.; et al. Large genomic differences between *Moraxella bovoculi* isolates acquired from the eyes of cattle with infectious bovine keratoconjunctivitis versus the deep nasopharynx of asymptomatic cattle. *Vet. Res.* **2016**, *47*, 31. [CrossRef] [PubMed]
35. Zoledowska, S.; Motyka-Pomagruk, A.; Sledz, W.; Mengoni, A.; Lojkowska, E. High genomic variability in the plant pathogenic bacterium *Pectobacterium parmentieri* deciphered from de novo assembled complete genomes. *BMC Genom.* **2018**, *19*, 751. [CrossRef] [PubMed]
36. Aylward, J.; Steenkamp, E.T.; Dreyer, L.L.; Roets, F.; Wingfield, B.D.; Wingfield, M.J. A plant pathology perspective of fungal genome sequencing. *IMA Fungus* **2017**, *8*, 1–15. [CrossRef] [PubMed]
37. A reference standard for genome biology. *Nat. Biotechnol.* **2018**, *36*, 1121. [CrossRef]
38. Kyaschenko, J.; Clemmensen, K.E.; Hagenbo, A.; Karltun, E.; Lindahl, B.D. Shift in fungal communities and associated enzyme activities along an age gradient of managed *Pinus sylvestris* stands. *ISME J.* **2017**, *11*, 863–874. [CrossRef]
39. Heeger, F.; Bourne, E.C.; Baschien, C.; Yurkov, A.; Bunk, B.; Sproer, C.; Overmann, J.; Mazzoni, C.J.; Monaghan, M.T. Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. *Mol. Ecol. Resour.* **2018**, *18*, 1500–1514. [CrossRef]
40. Kronenberg, Z.N.; Fiddes, I.T.; Gordon, D.; Murali, S.; Cantsilieris, S.; Meyerson, O.S.; Underwood, J.G.; Nelson, B.J.; Chaisson, M.J.P.; Dougherty, M.L.; et al. High-resolution comparative analysis of great ape genomes. *Science* **2018**, *360*. [CrossRef]

41. Sutton, J.T.; Helmkampf, M.; Steiner, C.C.; Bellinger, M.R.; Korlach, J.; Hall, R.; Baybayan, P.; Muehling, J.; Gu, J.; Kingan, S.; et al. A high-quality, long-read *de novo* genome assembly to aid conservation of Hawaii's last remaining crow species. *Genes* **2018**, *9*, 393. [CrossRef] [PubMed]
42. Ameur, A.; Kloosterman, W.P.; Hestand, M.S. Single-molecule sequencing: Towards clinical applications. *Trends Biotechnol.* **2019**, *37*, 72–85. [CrossRef] [PubMed]