Comparison of BGI and Illumina sequencing technologies for single-cell RNA sequencing

Anne Senabouth[1], Stacey B Andersen[2], Qianyu Shi[3], Lei Shi[3], Feng Jiang[3], Wenwei Zhang[4], Kristof Wing[5], Maciej Daniszewski[6,7,8], Samuel W Lukowski[2], Sandy SC Hung [8], Quan H Nguyen[2], Lynn Fink[9,10], Anthony Beckhouse[9], Alice Pébay[6,7,8], Alex W Hewitt[5,8], **Joseph E Powell[1, 11]**

[1]Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Darlinghurst, Sydney, [2]Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, [3]MGI, BGI-Shenzhen, Shenzhen 518083, China, [4]BGI-Shenzhen, Shenzhen 518083, China, [5]Menzies Institute for Medical Research, School of Medicine, University of Tasmania, Hobart, [6]Department of Anatomy and Neuroscience, The University of Melbourne, Parkville, Melbourne, [7]Department of Surgery, The University of Melbourne, Parkville, Melbourne, [8]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Melbourne, [9]BGI Australia, Brisbane, [10]Diamantina Institute, The University of Queensland, Woolloongabba, Brisbane, [11]St Vincent's Clinical School, University of New South Wales, Sydney

Introduction

- High-throughput single-cell RNA sequencing experiments require considerable amounts of sequencing to maximize capture and coverage individual cell transcriptomes [1]
- In droplet-based scRNA-seq, performance of sequencer affects the quantity and quality of the dataset due to the use of nucleotide barcode tags to identify cells and individual mRNA molecules [2]
- Majority of scRNA-seq datasets are generated with sequencing platforms developed by Illumina
  - Platforms include NextSeq 550, HiSeq X series, NovaSeq 6000
  - We compared the performance of the NextSeq 550 and NovaSeq 6000 sequencers to the MGISEQ-2000 sequencer developed by BGI

Methodology

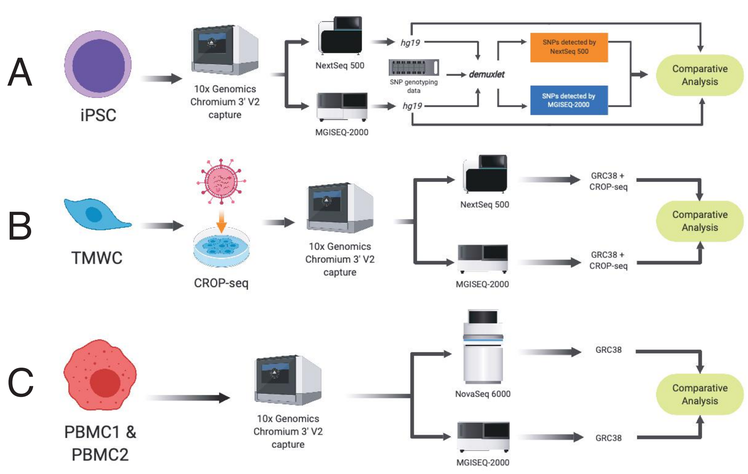

Figure 1. Preparation of single cell libraries for different experiment types and sequencing using Illumina and BGI platforms and subsequent analysis. A: Human **I**nduced **P**luripotent **S**tem **C**ells (**iPSC**) were generated from human donors and underwent genotyping in addition to pooled scRNA-seq. B: **T**rabecular **M**esh**w**ork **C**ells (**TMWC**) were derived from iPSCs and screened with a CRIPSR-based molecular screen (CROP-seq). C: **P**eripheral **B**lood **M**ononuclear **C**ells (**PBMC**) were extracted from patients, combined into two separate pools and sequenced using scRNA-seq.

1. Single cell libraries were generated from samples and sequenced as described in **Figure 1**.
2. Additional comparisons were performed on iPSC and TMWC experiments
   a) iPSC: Identification of genetic variation from multi-donor single cell pool
   b) TMWC: Detection of guide RNA (gRNA) in cells

Results

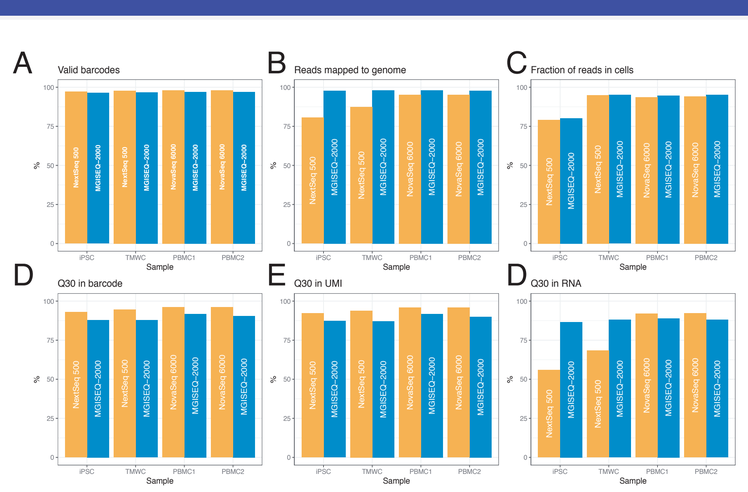

Figure 2. Summary of key metrics related to sequencing and mapping. A: Fraction of reads associated with a valid barcode. B: Fraction of reads that mapped to a location in the reference genome. C: Fraction of reads that are associated with a valid cell barcode, and has been mapped to the genome. D, E: Fraction of bases associated with a cell/UMI barcode with a Q-score of 30 or greater. F: Fraction of bases associated with cDNA with a Q-score of 30 or greater.

Results

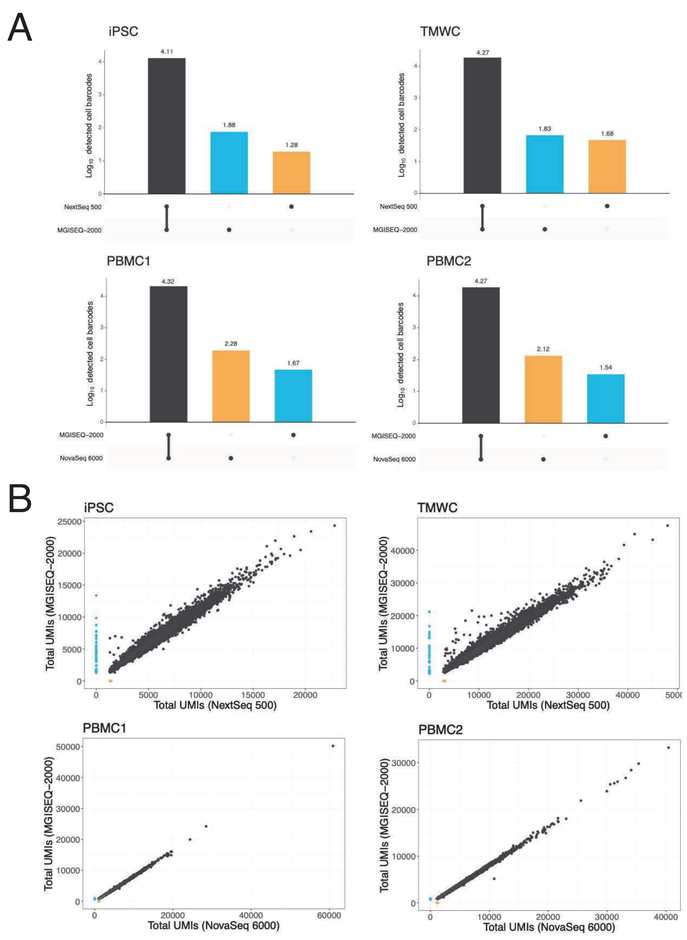

Figure 3. Cell barcodes and UMIs detected by platforms. A: UpSets represent $\log_{10}$ number of cell barcodes detected that were exclusive, and common to both platforms. B: Scatter plots represent the relationship between the total UMI content of a cell detected by both platforms.

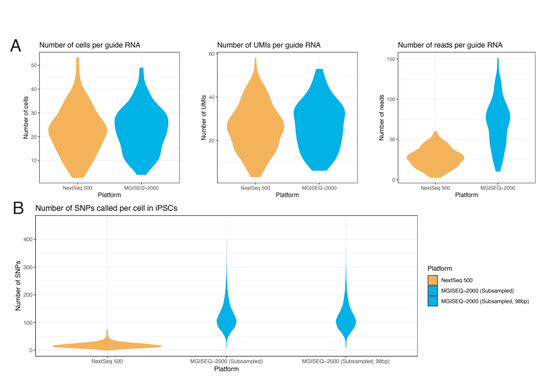

Figure 5. Comparison of metrics for CROP-seq and variant calling experiments. A: Definitive assignment of guide RNA to UMIs and cells, and number of reads per guide RNA. B: Distribution of SNPs per cell, by platform.

Results

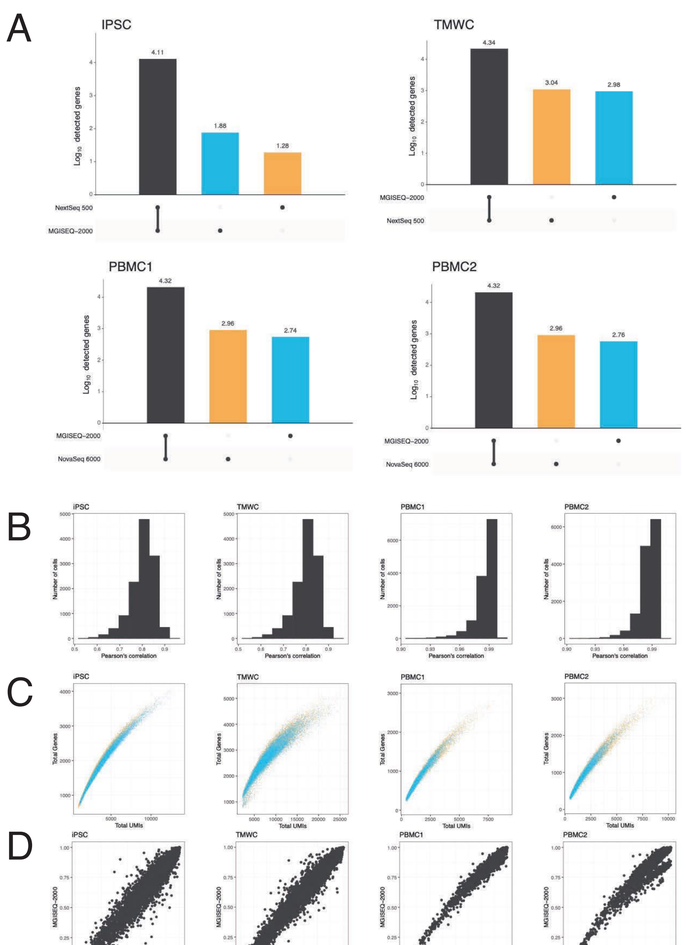

Figure 4. Gene expression profiled by platforms. A: UpSets represent $\log_{10}$ number of genes detected that were exclusive, and common to both platforms. B: Correlation of gene expression between cells common to both platforms C: Capture efficiency between cells sequenced by both platforms. D: Correlation of dropout rates calculated by M3Drop.

Prediction	NextSeq 500	MGISEQ-2000	MGISEQ-2000 (98bp)
Unassigned	6016	4322	4311
Correctly assigned	4272	5966	5977

Table 1. Summary of donor assignments by *demuxlet*. Variants were called from 306,414 markers to deconvolute a pool of cells containing iPSCs derived from two individuals.

Discussion

- MGISEQ-2000 and NovaSeq 6000 are highly comparable in performance and generated data, based on sequencing quality (**Figure 2**), cell **(Figure 3),** molecule and gene detection (**Figure 4**).
- MGISEQ-2000 performed better than the NextSeq 550 in cell, molecule and gene detection
- A higher proportion of bases generated by the MGISEQ-2000 that were associated with barcode sequences had a Q-score that was less than 30, in comparison to reads generated by Illumina sequencers **(Figure 1)**
  - More investigation is required to determine the cause of this difference in quality
  - Impact of lower Q30 in cell barcodes are mitigated by the correction step in the Cell Ranger pipeline
- NextSeq 550 and MGISEQ-2000 detect guide RNA at similar frequencies (**Figure 5**)
- Demuxlet called 1,065,659 more SNPs in reads generated by the MGISEQ-2000
  - Demuxlet was able to assign more cells to their donor with MGISEQ-2000 data (**Table 1**)
  - Difference is due to the MGISEQ-2000 producing a greater proportion of quality RNA reads (**Figure 2**)

Conclusion

- The performance of the MGISEQ-2000 for scRNA-seq experiments is highly comparable to the NovaSeq 6000
- Even after read-depth equalization, the MGISEQ-2000 performs better than the NextSeq 500 in the detection of cells, genes and UMIs
- There was no difference in guide RNA detection for CROP-seq experiments
- The MGISEQ-2000 detected more SNPs than the NextSeq 500.

References

1. Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Medicine, 9(1). https://doi.org/10.1186/s13073-017-0467-4
2. 10x Genomics (2017), *Base Composition of Sequencing Reads of Chromium™ Single Cell 3' v2 Libraries*. CG00080 Rev B Technical Note. Available at https://support.10xgenomics.com/single-cell-gene-expression/sequencing/doc/technical-note-base-composition-of-sequencing-reads-of-chromium-single-cell-3-v2-libraries

Contact

Joseph Powell
- Email: j.powell@garvan.org.au
- Twitter: @drjosephpowell

Anne Senabouth
- Email: a.senabouth@garvan.org.au
- Twitter: @senabouth