Evaluation of three next generation sequencing platforms for immune repertoire sequencing

Chunlin Wang[1], Shawn Levy[2,] Qunying Yang[1,2], Miranda Byrne-Steele[1,2], Jian Han[1,2]
[1]iRepertoire Inc., [2]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

Abstract

The advent of next generation sequencing (NGS) techniques, combined with a semi-quantitative multiplex PCR method, enables us to comprehensively profile the lymphocyte receptor diversity (immune repertoire) of the entire collection of B or T cells from a particular sample. Different NGS platforms vary in read length, throughput, sequencing error rate and error pattern, hands-on time, turnaround time and price. Three popular NGS techniques (Illumina HiSeq, 454 FLX and Ion Torrent) were employed to sequence amplicons from T cell receptor clones generated with arm-PCR in order to evaluate the usage of those techniques on immune repertoire sequencing. The error pattern and rate of the three techniques in the context of T cell receptor sequences were profiled. Based on the differences of the three techniques, we discuss the usage of each NGS platform in the context of immune repertoire sequencing.

Sequencing length

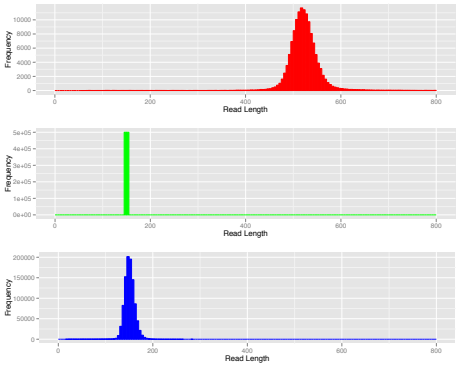

Figure 1. The distribution of read length of three popular NGS platforms for 454 GS Junior (top), Illumina MiSeq (middle), and Ion Torrent (bottom). The gap between FR1 and CDR3 is about 240 bp and between FR2 and CDR3 is about 140 bp. The length of the J domain is around 50 bp. Taking all of this information into consideration, the 454 technology allows the sequencing of V-D-J molecules from FR1 region up to the C region, while the other two technologies are useful for sequencing V-D-J molecules from FR3 region to the C region. For those studies where hypermutation information is essential, the 454 technology is the only choice currently.

Error pattern

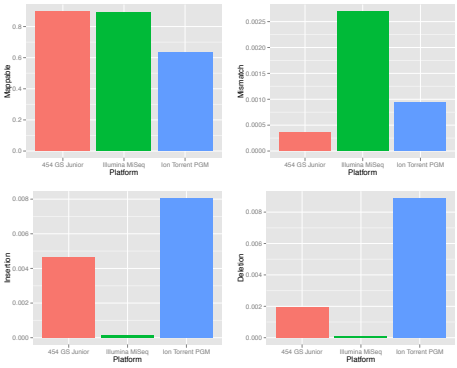

Figure 2. Evaluation of error pattern of three platforms: 454 GS Junior (red), Illumina MiSeq (green), and Ion Torrent PGM (blue) in terms of the percentage of mappable bases (top left), mismatch error (top right), insertion error (bottom left), and deletion (bottom right). Those numbers are based on re-evaluation of sequence data with the access numbers (SRR388806 for 454 GS Junior, SRR389193 for Ion Torrent PGM and SRR390202 for Illumina MiSeq). Among the three platforms, PGM has the least percentage of mappable bases. After soft clip, the mismatch error rate of MiSeq is about three and six times greater than that of PGM and that of 454, respectively. On the other hand, the insertion/deletion error rate of MiSeq is the lowest among three platforms. The error rate around homopolymeric region of both 454 and PGM reads is much higher than regular regions[1].

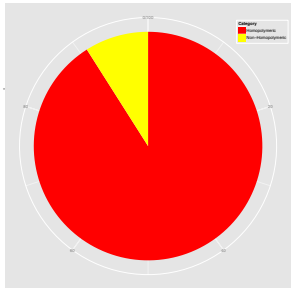

Figure 3. Distribution of CDR3 with or without homopolymeric regions. Homopolymeric regions are defined as a stretch of 3 or more identical bases.

Sequencing VDJ with three Tech.

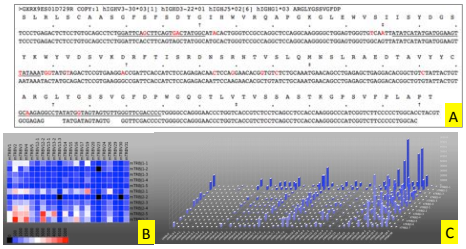

Figure 3. Example of applying three sequencing platforms (A: 454 FLX, B:Ion Torrent and C:Illumina HiSeq2000). In panel A, red color highlights hypermutation bases and underline indicates the CDR1, CDR2 and CDR3.

Conclusions

•Three popular NGS platforms were evaluated for sequencing VDJ molecules. Among the three, the 454 technology offers the longest reads, which could cover the entire VDJ region and allow for profiling of the hypermutation pattern, an important parameter for Ig molecules. The other two technologies produce reads under 200 bp, which limit their usage to profiling CDR3 distribution only.

•Among the three platforms, PGM has the least mappable bases and the highest insertion/deletion error rates. After soft clipping, MiSeq has the highest mismatch error rates. More errors were found in the homopolymeric region of 454 and PGM reads, which might limit the usage of 454 and PGM platform on VDJ sequencing as around 90% of CDR3s have at least one homopolymeric region.

•Our experience of VDJ sequencing[2] with three platforms suggests that Illumina technology, HiSeq2000 in particular, due to its high throughput and when used in combination with bi-directional sequencing, allows profiling of the CDR3 distribution with deep coverage. On the other hand, the 454 technology allows profiling the hypermuation pattern of Ig molecules.

References

1. Nicholas J Loman et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012 Apr 22. doi: 10.1038.

2. Wang, Chunlin *et al.* High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *PNAS.* 107,1518-1523.(2010).