600 base reads on the Ion S5™ Next-Generation Sequencing System enables accurate HLA typing of 96 samples on one 530™ chip.
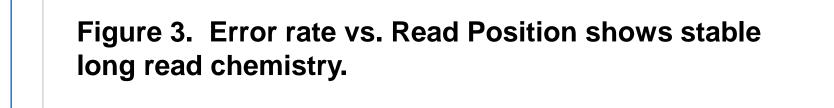
Peter B. Vander Horn, Cisilya Duncan, Jamsheed Ghadiri, Amneet Gulati, Diana Jeon, April Jung, Mindy Landes, Tommie Lincecum, Geoffrey Lowman, Vadim Mozhayskiy, Linus Ong, Xinzhan Peng, Maryam Shenasa, Prasanna Thwar.
Thermo Fisher Scientific, 5791 Van Allen Way, Carlsbad, CA, USA, 92008

ABSTRACT

Longer read lengths simplify genome assembly, haplotyping, metagenomics, and the design of library primers for targeted resequencing. Several new technologies were developed to enable the sequencing of templates with inserts over 600 bases: a fast isothermal templating technology, an ISP™ that is optimized for maximum template density, a new long-read sequencing polymerase, and instrument scripts that consume less reagents. We demonstrate the combination of these technologies to sequence 600 base long DNAs on an Ion 530 Chip™ with an average AQ20 mean read length over 500 bp. The protocol was used to type human leukocyte antigen (HLA) alleles, a haplotyping application that is greatly simplified by long read length sequence data. 96 HLA samples were typed with 99.7% concordance to truth on one Ion 530 chip.
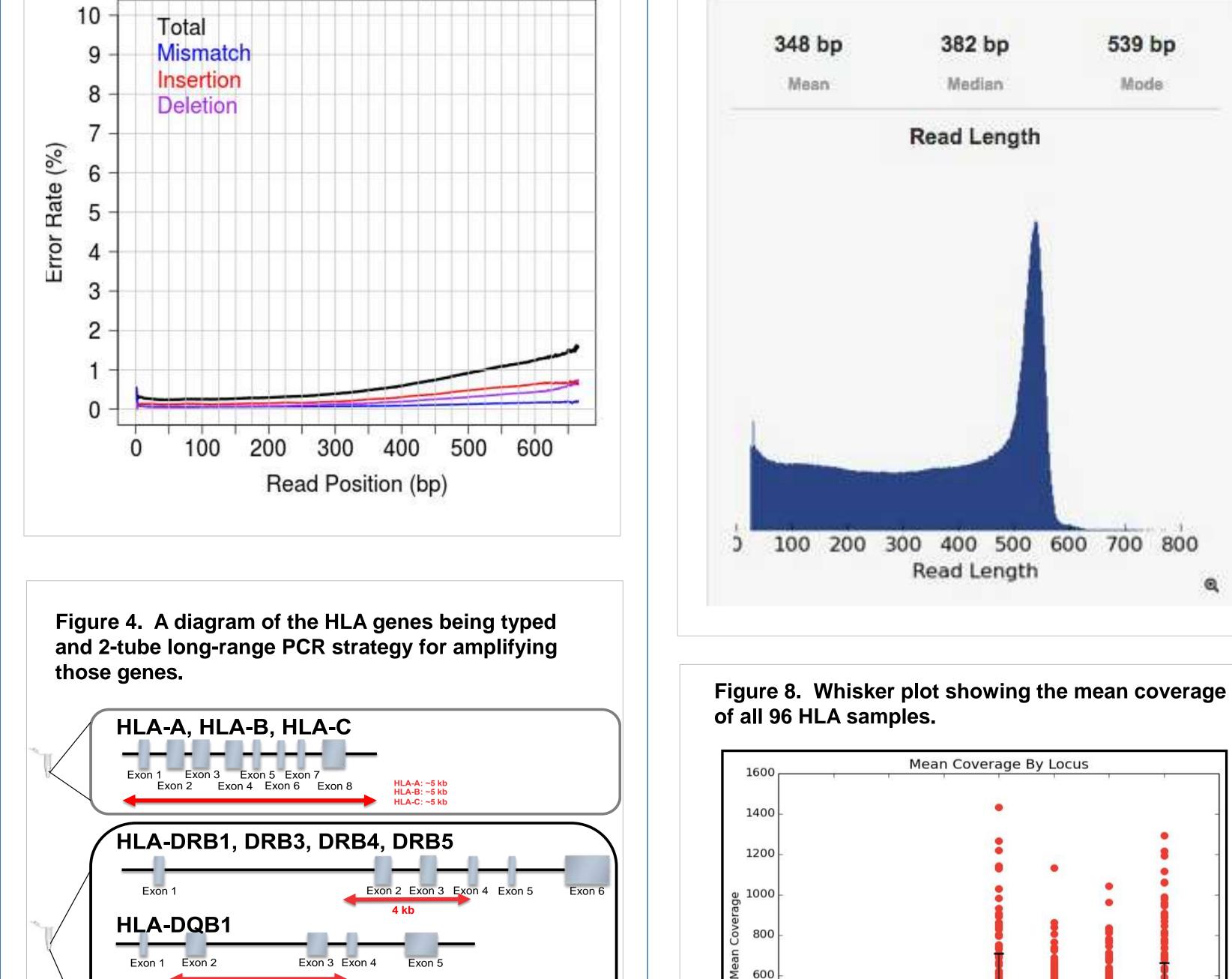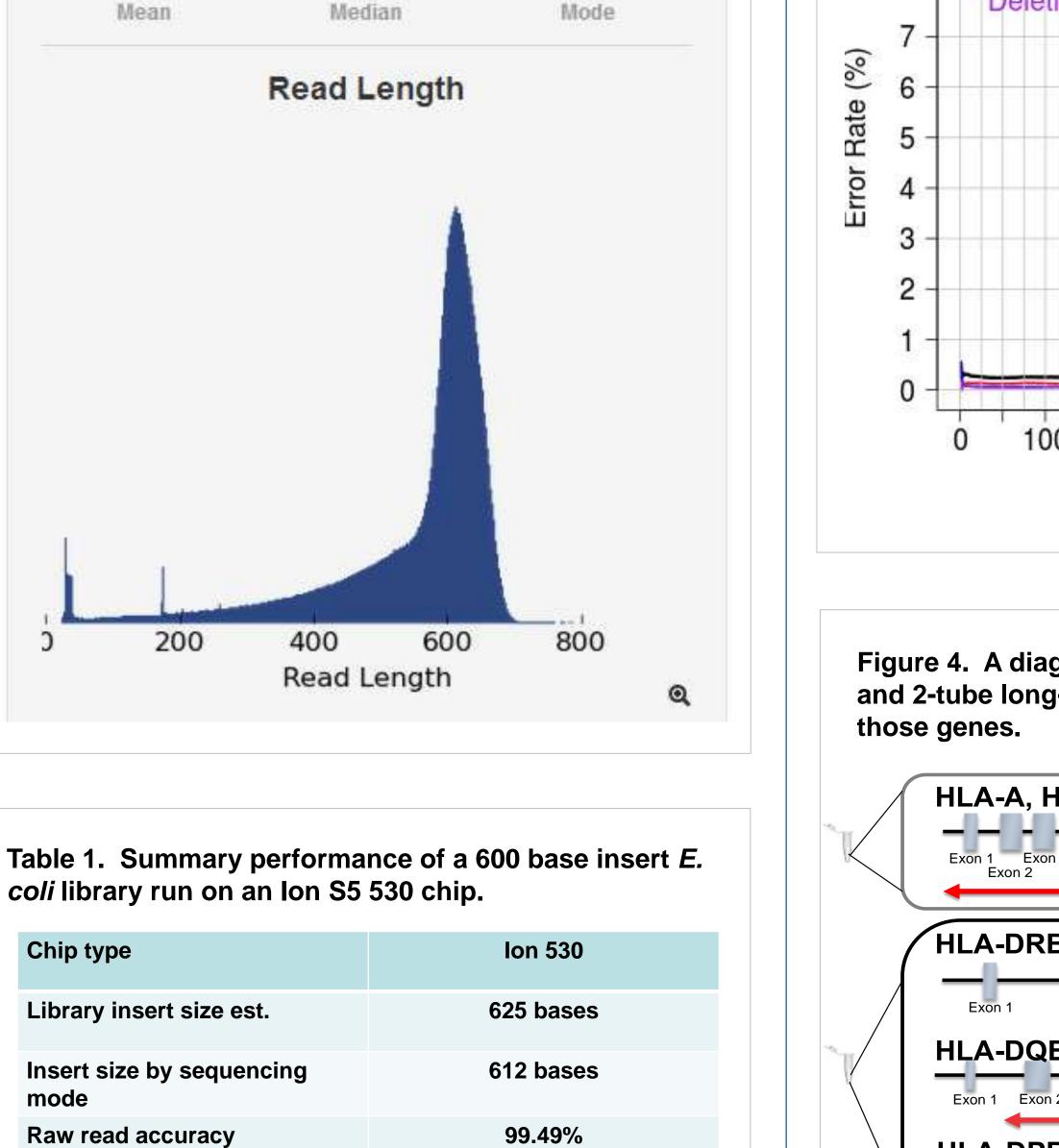
INTRODUCTION

Several applications, including HLA typing and immune repertoire (1) are better enabled by 600 base long reads. HLA is a particularly challenging application because it requires haplotyping epitopes that are relatively far from each other. Long reads are particularly helpful for maintaining the phase of polymorphisms in this rapidly evolving part of the human genome.

Thermo Fisher's RUO NXType™Kit is a research use only solution for HLA typing by NGS. The system features long range primers and an improved chemistry for use on the Ion Torrent Personal Genome Machine (PGM)™, enabling sequencing of HLA Class I & II genes in three days (Fig 5). The NXType system includes: a comprehensive kit containing HLA Primer Mixes, reagents and HLA TypeStream™, One Lambda's NGS Analysis Software. We sought to improve the NXType workflow and throughput by adapting it to the S5 NGS system, whose 530 chip offers ~ 4 times the throughput of the PGM 318 chip.

We have developed a number of sequencing biochemistry improvements to support accurate sequencing of 600 base inserts. The IA reaction was adapted to increase speed, yield, and longer templates. New Ion sequencing beads were developed with the optimal size, density, and primer load to accommodate longer templates on the 530 chip. A new sequencing polymerase improved signal and phase. And lastly, high-efficiency S5 instrument scripts were written to conserve reagents and increase the number of nucleotide flows; resulting in > 600 bases read lengths on a 530 chip run on the S5 platform. We then show that we can apply these improvements to type 96 HLA samples on one 530 chip with 99.7% concordance.

MATERIALS AND METHODS

Templates were produced by fragmenting to an average size of 600-700bp using Ion Shear™ enzymatic fragmentation technology. Ligation and nick-repair were performed using Ion Xpress adapters. The Pippin Prep system was used for size selection using 1.5% Dye free gels (Sage Science, CDF1510). Libraries were PCR amplified and purified with Agencourt AMPure XP beads. Amplified library yields were determined using a Qubit fluorometer and size distribution (Fig1) was determined using an Agilent 2100 Bioanalyzer™ (Agilent Technologies, G2938C) with the Agilent High Sensitivity DNA Kit (Agilent Technologies, 5067-4626). A.

DNAs were templated to Ion Sphere Particles (ISPs) using an isothermal amplification biochemistry that enables efficient templating of long templates onto ISPs. The ISP utilized was the result of a project to optimize ISPs for size, density, and primer concentration for maximal loading of long templates. The samples were sequenced using a sequencing polymerase that was evolved for producing longer reads on semiconductor chips. The sequencer was a commercially released S5 running scripts that rationed reagent consumption for maximum flows.

Figure 1. Bioanalyzer traces of *E. coli* DNAs are 717 bases. Ion adaptors total 70 bases in length

Figure 2. Read length histogram produced by sequencing a ~600 base-insert, *Escherichia coli* genomic library on an S5 530 chip.

549 bp	599 bp	612 bp
Mean	Median	Mode

Read Length

Table 1. Summary performance of a 600 base insert *E. coli* library run on an Ion S5 530 chip.

Chip type	Ion 530
Library insert size est.	625 bases
Insert size by sequencing mode	612 bases
Raw read accuracy	99.49%
Error @ 600 bp	1.24%
Error @ 400 bp	.602%
Reads (M)	15
Throughput	7.84 Gb

RESULTS

The insert size of the *E. coli* library was estimated to be between 600 and 650 bases based upon capillary chip analysis (Fig. 1). The library was templated and sequenced on an S5 Ion 530 chip using a sequencing biochemistry designed for long accurate reads.

The S5 System produced a read length histogram with a mode at 612 bases (Fig 2) and a raw read accuracy of 99.5% (Table 1). There were over 15 million reads with an average AQ20 mean read length of 537 bases. Total throughput was over 7.8 G of AQ20 bases. Interestingly the read quality remained quite high the entire length of the read (Fig. 3). Raw read accuracy was 98.75% at the 600th base.

We prepared 96 HLA samples provided by our sister company, TDx (formally One Lambda) as specified in the materials and methods section. The HLAs allele had been determined previously for these samples using sequenced based typing (SBT). The samples were combined after barcodes were attached. Bioanalyzer results indicated that gel size selection resulted in an insert around 550 bases, excluding 20 base long barcodes. The sample was templated in one IA reacting and run on one S5 530 chip.

The result was a read length histogram with a mode at 539 bases. Sequencing accuracy metrics are not informative for this rapidly evolving region of the human genome. The concordance rate between the SBT-derived genotypes and those generated by TypeStream auto-analysis was 99.7% for the HLA-A, B, C, DRB1, DQB1, DPB1 and DBR 345 loci combined. Thus, accurate HLA typing on 96 samples was demonstrated on 1 530 chip.

Ambiguous allele calls were found to be due to low coverage. Examining the coverage of each of alleles (Fig 8.) we identified that the ambiguous calls were due to dangerously low coverage of HLA-C. (Default parameters require 20 reads to make a call.) Average coverage depth for all of the class I genes was between 100 and 200, compared to between 400 and 600 for the class II genes. This result suggests that with some slight adjustments in the long PCR input, the number of alleles assayed on a 530 chip might be comfortably doubled to 192 samples per chip.

Figure 3. Error rate vs. Read Position shows stable long read chemistry.

Figure 4. A diagram of the HLA genes being typed and 2-tube long-range PCR strategy for amplifying those genes.

Figure 5. An introduction to the NXType™ work flow.
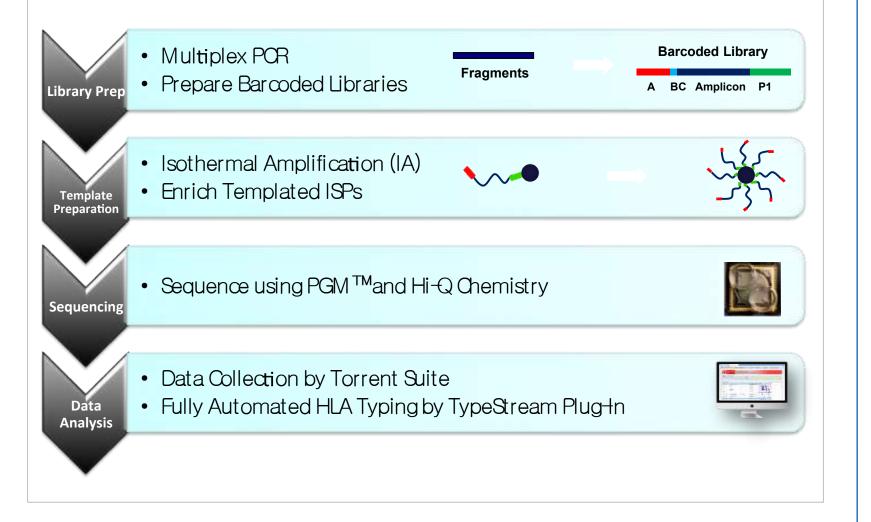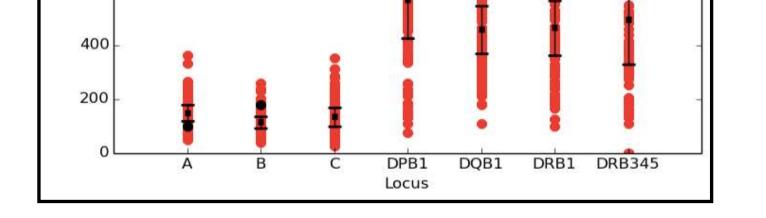
Figure 6. A screen shot from HLA TypeStream™, TDx's NGS Analysis Software showing intuitive data display.

Table 2. Metrics of 96 HLA samples run on an Ion S5 530 chip.

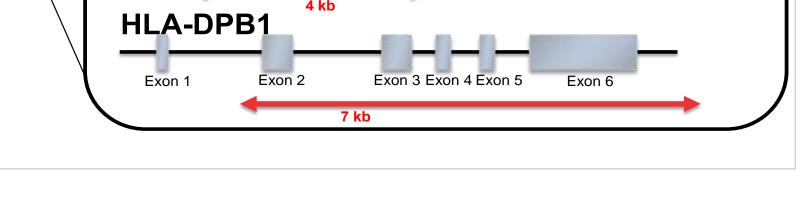| Chip type	Ion 530
Library insert est.	550 bases w/o barcode
Read length by sequencing mode	539 base mode
Reads (M)	16.7

Figure 7. Read length histograms of 96 barcoded HLA samples run on an Ion S5 530 chip.

348 bp	382 bp	539 bp
Mean	Median	Mode

Read Length

Figure 8. Whisker plot showing the mean coverage of all 96 HLA samples.

CONCLUSIONS

We have demonstrated that by combining improvements in templating and sequencing biochemistry we are able to sequence templates longer than 600 bases with high accuracy on an Ion S5 530 chip.
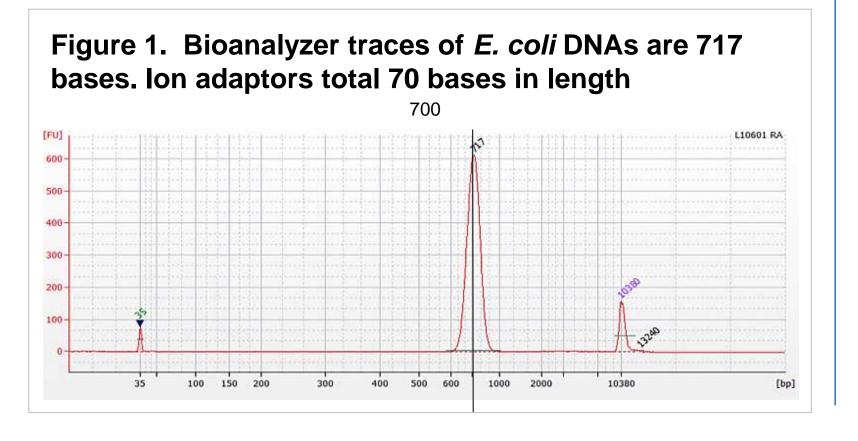
These improvements open the S5 use space to include haplotyping applications that require longer reads. As a demonstration of that, we accurately typed 96 HLA samples on one 530 chip.

REFERENCES

1. J.J.J.A Calis and B.R. Rosenberg, Characterizing Immune Repertoires by High Throughput Sequencing: Stregeties and Applications. Trends in Immunology, Vol. 35, 581-590 (2014)
2. J.C. Barone, K. Saito, K. Beutner, M. Campo, W. Dong, C.P. Goswami, E.S. Johnson, Z-X. Wang, S. Hsu. HLA-genbotyping of Clinical Specimins using Ion Torrent-Based NGS., *Human Immunology* Vol 76, 903-909, 2015
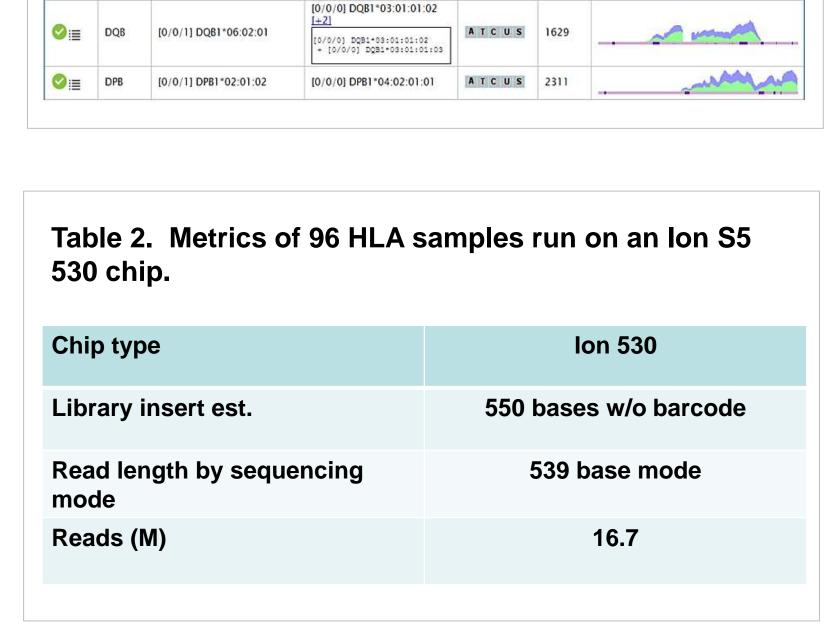
ACKNOWLEDGEMENTS

For Research Use Only. Not for use in diagnostic procedures.

Thermo Fisher
SCIENTIFIC