# Simultaneous Evaluation of Small RNA, Whole Transcriptome, Whole Genome, ChipSeq and Targeted Resequencing on Next Generation Sequencing Platforms

N. Patel, N. Fong, M. Chan, D. Cox, D. Hom, C. Inman, P. Kinney, A. Kivett, R. Padilla, M. Matvienko, A. Surdutovich, M. Storm, T. Burcham, H. Shepherd, L. Joe
Life Technologies, 850 Lincoln Centre Drive, Foster City, CA, USA, 94404

## ABSTRACT

Researchers are using Next Generation Sequencing platforms to support more and more applications. It is critical for such platforms to be flexible enough to support many applications at the same time while maintaining low cost, high quality and throughput. The 5500 Series Genetic Analyzers are able to interrogate up to twelve independent samples or applications run simultaneously. In a single run we sequenced a single-plex human whole transcriptome sample, a multi-plex of universal human reference, human placenta, human brain reference and HeLa whole transcriptome sample, a microbial whole genome sample, and a human placenta smRNA sample. Each sample type used required a different sequencing run, yet with this new process they all were run together saving time, reagents, and freeing the researcher from having to wait to collect sample of the same type to run together. In this study, we describe the development of simultaneously sequencing 6 samples in a single run.

## INTRODUCTION

Next-generation sequencing technology is quickly being utilized in many applications for discovery and clinical research. The applications are very diverse, from DNA and RNA sequencing to small target region resequencing to whole genome resequencing. In basic and clinical research, especially in small core and laboratories, an experiment may have many samples, but a small target genomic/gene region. Alternatively, some labs may have very few samples but require large amounts of data and high throughput. We present the 5500 and 5500xl Genetic Analyzers as the latest addition to the Life Technologies Genetic Analysis sequencing platforms. The 5500 Series Genetic Analyzers uses the fundamentals of the SOLiD™ sequencing technology and integrates new features to make a more economical and robust instrument possible. One of the newest components of the instrument is the 6-Lane FlowChip, a micro-fluidic device that enables a operators to configure each lane for active/non active, chemistry type Fragment/Paired End/Matepair and readlength. The reagents are consumed only for active lanes--enabling the Pay-Per-Lane sequencing feature. Each lane in the FlowChip to be individually configurable for independent applications and readlength, thereby eliminating reagent waste and reducing wait time for assembling samples of the same application. We present the results of simultaneously running all six lanes on one FlowChip with 5 unique samples on the 5500 Genetic Analyzer.

## MATERIALS AND METHODS

5500 Genetic Analysis Instrument
Sequencing Reagents/Buffers
Deposition Kit + FlowChip
DNA Samples listed in Table 1
LifeScope Secondary Analysis Software

**Table 1. FlowChip Lane Description**

| Lane | Library | Run Length |
|---|---|---|
| 1 | NimbleGen VCRome 2.0.0.1 Exome capture | 1x50 +ECC |
| 2 | Kinome Enrichment Study | 1x50 +ECC |
| 3 | MCF7 Whole Transcriptome | 1x50 +ECC |
| 4 | Chromatin ImmunoPrecipitation (ChIP-Seq) | 1x35 +ECC |
| 5 | Human Reference Long Mate Pair (HuRef LMP) | 2x50 +ECC |
| 6 | | |

Table 1- describes each sample deposited in the FlowChip and number of cycles run for each lane.

## RESULTS

**Figure 1. 5500 Genetic Analysis Systems**

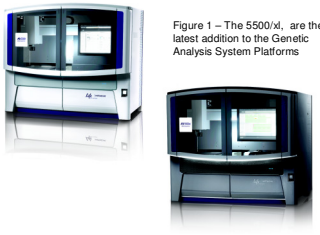


Figure 1 – The 5500/xl, are the latest addition to the Genetic Analysis System Platforms

**Figure 2. Sequencing Tag Run Order**

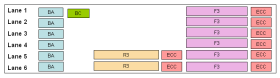



Figure 2 - This is a representation of the run schedule for the flowchip. It demonstrates the order in which the sequencing run will take place. A BA is a bead assessment that evaluates the density of the lane, and determines the position of each sample and sequencing control bead. Once the density of each lane has been determined the next step to run all samples that have a Barcode (BC). Followed by Tag 2 (R3) sequencing, then Tag (F3) sequencing. All Exact Color Call Sequencing is done after the tag has been sequenced. The Exact Call Chemistry enables the translation from color space into base space.

**Figure 3. Deposited FlowChip**

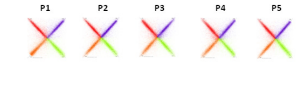
Figure 3 – The FlowChip used for this test is figured above, the image was taken after the sequencing run completed. All 6 lanes were deposited at the same time and run according to the Tag run schedule from Figure 1. Some bead loss did occur during the run, however it did not affect the quality of sequencing data.

**Table 2. Lane Density and Estimated Throughput**

| Lane | Density Beads / Panel | Estimated Throughput (bases) |
|---|---|---|
| 1 | 238K | 6,896,071,250 |
| 2 | 137K | 4,308,669,300 |
| 3 | 139K | 4,152,333,050 |
| 4 | 200K | 4,093,611,585 |
| 5 | 235K | 6,584,518,750 |
| 6 | 241K | 6,709,807,200 |

Table 2 - Each lane was deposited at a target density of 250K/panel, however depending upon bead concentration and quality of beads the actual density achieved can vary. These densities were obtained after running the Bead Assessment sequencing portion of the run was completed. Based on those densities the throughput of the entire run based on cycles can be estimated.

**Figure 4. Barcode Distribution**



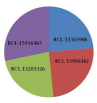



Figure 4 – After Bead Assessment, Barcoding is the second step in a sequencing run. When sequencing barcodes, it is important to have an equal distribution of barcode tags. The satay plot shows good sequencing of each of the barcode tags (5 basepairs) and the pie chart shows an even distribution of barcode reads from secondary analysis.

**Figure 5. Barcode Detection and Mapping**

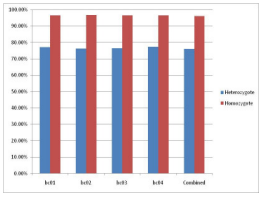


Figure 5 - Lane 1 consisted of a Targeted Resequencing sample with 4 barcodes. The sequencing run for this lane showed that SNP concordance is good and consistent across the 4 samples even at low coverage.

**Table 3. Long Mate Pair Sequencing Mapping Statistics**

| Lanes | Sample | Library | Total Reads | Mapped Reads |
|---|---|---|---|---|
| 5 | HuRef LMP | 2x50 +ECC | 131,690,375 | F3: 85.18% R3: 71.80% |
| 6 | | | 134,196,144 | F3: 82.07% R3: 76.16% |

Table 3 - The 3rd step in the sequencing run is to run any lanes that have 2 tags. Lanes 5 and 6 both contained the same LMP sample and in during the 3rd step tag 2 was sequenced. The analysis of the 2 lanes showed similar secondary sequencing analysis.

**Table 4. Forward Sequencing Mapping Statistics**

| Lane | Total Reads | Mapped Reads |
|---|---|---|
| 1 | 126,174,252 | 92.88% |
| 2 | 86,173,386 | 91.49% |
| 3 | 83,046,661 | 84.10% |
| 4 | 117,006,936 | 85.28% |

Table 4 - The fourth step in the sequencing run is to sequence the forward tag (DNA closest to the bead). All lanes are run simultaneously during this step of the run. Lane 5 and 6 data is shown in Table 3.

**Table 5. Lane 2 Kinome Enrichment Sequencing Coverage**

| | |
|---|---|
| Average depth of coverage within targets | 688.53 |
| Percent of target bp covered at ≥ 1X: | 95.57% |
| Percent of target bp covered at ≥ 5X: | 92.59% |
| Percent of target bp covered at ≥ 10X: | 90.62% |
| Percent of target bp covered at ≥ 20X: | 87.87% |
| Percent of target bp not covered: | 4.4% |

Table 5 - Lane 2 consisted of a Kinome Study that ran for a total of 60 cycles (50 cycles + ECC). Using LifeScope we can determine the percentage of sequenced information for a variety of coverage levels ranging from 1X – 20X coverage.

**Figure 6. SNP concordance**

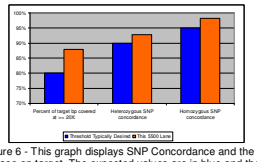



Figure 6 - This graph displays SNP Concordance and the percentage of base on target. The expected values are in blue and the values obtained from this sequencing run are in orange. For all analyzed metrics the 5500 Genetic Analyzer has performed at levels greater than expected.

**Figure 7. Motif Detection of ChipSeq Sample**

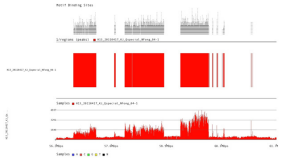


Figure 7 – Lane 4 consisted of a ChipSeq sample that contained known motifs, which were detected by the 5500 Genetic Analyzer.

**Figure 8. Whole Transcriptome SOLiD™ vs. 5500 Genetic Analyzer Data Comparison**

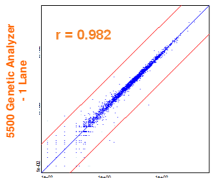



Figure 8 – The Whole Transcriptome sample run in Lane 3 was also run on a SOLiD™ Instrument, the previous sequencing platform. There is a high correlation in data between the 2 platforms, showing that the 5500 Genetic Analyzer matches performance of our previous platform when using the same sample.

**Figure 9. Sample Quality Value Distribution**

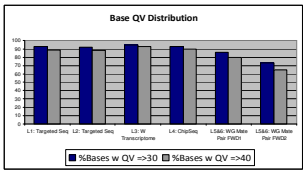



Figure 9 – The percentage of bases with quality values greater than 30 and 40.

## CONCLUSIONS

The 5500 Series Instruments are the latest addition to the Life Technologies Genetic Analysis Sequencing platform. The simultaneous running allows users to save money and time by setting up the various lanes for different sequencing type runs. Which also add benefits to reduce wait time to run only 1 sample type. Now users can run a variety of sample types on the same chip and still achieve great data performance and reduced sequencing costs. As each run is completed secondary analysis of the sample can be started.

In our study we were able to show that you can run the same sample across multiple lanes along with other types of sample in the other lanes and achieve data quality that is similar to and better than the previous platform. For all lanes we were able to obtained % mapped reads greater than expected. Our control samples also indicated the sequencing run performed well because we were able to detect all known motifs (lane 4) and our data analysis highly correlated to a sequencing run on SOLiD™ using the same sample type.

## ACKNOWLEDGEMENTS

Thank you to everyone working on the 5500 Genetic Analysis System at Life Technologies. Special Thanks to Rachel Fish, Janet Berena, Catalin Barbacioru, Yongming Sun, Matt Dyer, Robert Nutter, Michael Rhodes and Quynh Doan for helping run, analyze and obtain the data for this presentation.

## TRADEMARKS/LICENSING