The High Throughput Genomic Sequencing Facility (HTSF) in the Integrated Genomics Cores is a full-service sequencing facility in the School of Medicine. It supports multiple sequencing platforms and state-of-the-art techniques to assist UNC investigators with genetic and genomic research. This facility has 2,500 sq. ft. of dedicated wet bench lab space housed in a new 9,000 sq. ft. facility along with the UNC gene expression core.

The HTSF has 1 NovaSeq 6000, 2 HiSeq 4000 and 2 HiSeq 2500 (Illumina), 4 MiSeq, GridION Oxford Nanopore System, as well as associated equipment necessary for efficient operation, including cBot cluster generators (Illumina), a capillary electrophoresis systems – 2 LabChipGX (Perkin Elmer), 2 Tape Stations (Agilent), 2 Bioanalyzes (Agilent); DNA shearing devices Covaris LE220, S2, and a host of other equipment including a Chromium system (10x Genomics) for single cell analysis and Saphyr (BionanoGenomics) for optical mapping of the genome. All sequencers are associated with on-board computers for real-time data processing, with one server dedicated for data analysis.
To increase efficiency and reproducibility of library preparation and pooling the facility is utilizing six robotic systems supporting unique and overlapping applications—two Genesis systems 150 and 200 (TECAN), one Biomek (Beckman Coulter), one Bravo (Agilent), one Sciclone G3 (Perkin Elmer) and Mantis (Formulatrix). Library preparation process for RNAseq application can be performed on more than one system to assure redundancy in our workflow.

Twenty-two full-time personnel, in addition to the Managing Director, Dr. Piotr Mieczkowski, staff the HTSF. Sequencing data processing, management, initial analysis and distribution of all data generated at the HTSF are handled by the UNC Center for Bioinformatics, are responsible for data processing, data management, initial analysis and distribution of all data generated at UNC HTSF. They have worked with a number of open-source software tools (including but not limited to SAMTOOLS, Bowtie, TOPHAT, BWA, SOAP2, MAQ etc.) that are used for analysis of deep-sequencing data. Hardware support for the facility includes four Dell Poweredge 2950 servers running Redhat enterprise Linux (EL4) (each with 2 x 2.66 GHz Quad core Xeon CPU’s, 32 GB RAM, 6 TB of disk storage) available for management/analysis of data from the next-gen sequencers within the Center, and a 30TB iSCSI SAN for next-gen sequence data analysis and storage. A small Linux cluster (20 node) was recently installed shortly to help facilitate large-scale analysis of sequence data. In addition to open source software programs meant for analysis of next-gen sequence data (maq, ChIP-seq, mosaik, phred/phrap/consed) there are two network licenses for CLC Genomics Workbench. Dr. Tristan De Buysscher, a senior bioinformatics scientist in the Center, provides dedicated programming support for informatics for high throughput sequencing projects.

***\*\*Please see the longer version supplied from the HTSF below\*\****

Next Generation Sequencing

* UNC-Chapel Hill maintains several core facilities focused on genome sciences. Part of UNC’s success in maintaining an exciting research atmosphere is its open-door policy for core research facilities. Most importantly for this project, our lab has extensive positive collaborations with the UNC High-Throughput Sequencing Facility (HTSF, http://www.unc.edu/htsf/), which is supported by the School of Medicine and the Cancer Center. They operate state-of-the-art next generation sequencing instruments. The HTSF has 1 NovaSeq 6000, 2 HiSeq 4000 and 2 HiSeq 2500 (Illumina), 4 MiSeq, GridION Oxford Nanopore System, as well as associated equipment necessary for efficient operation, including cBot cluster generators (Illumina), a capillary electrophoresis systems – 2 LabChipGX (Perkin Elmer), 2 Tape Stations (Agilent), 2 Bioanalyzes (Agilent); DNA shearing devices Covaris LE220, S2, and a host of other equipment including a Chromium system (10x Genomics) for single cell analysis and Saphyr (BionanoGenomics) for optical mapping of the genome.
* HiSeq2500 system is a gold standard in RNAseq sequencing since it was used as a main data production system for The Cancer Genome Atlas project (TCGA). This system is using 4 color chemistry (each of the nucleotides has assigned one color) and random cluster generation flowcells. In our production facility this system is able to produce 32 lanes of sequencing per week with capacity around 230 million clusters per lane. Although it is older Illumina technology it still very popular among investigators interested in replication of TCGA RNAseq workflow for non-bias data comparison.
* HiSeq4000 system is currently workhorse for many sequencing applications including RNAseq. This system leverages innovative patterned flow cell technology to provide rapid, high-performance sequencing using 4 color chemistry. Perform production-scale, high-throughput exome or transcriptome sequencing projects quickly and economically. It can deliver 48 lanes of sequencing in 9 days and each lane can produce sequencing data from around 320 million clusters.
* NovaSeq 6000 is the state-of-the-art sequencing system from Illumina. It combines patterned large scale flowcells with new 2 color chemistry (combination of colors for each of 4 nucleotides). It can provide maximum sequencing output of the 20 billion clusters in 4 days. Output depends on a type of the flowcell and can be used for both medium and large-scale sequencing projects. Our tests indicate no technology bias for RNAseq comparing to the HiSeq4000 system. Therefore, we can predict that this technology soon become dominant for counting applications as well as resequencing of the genomes.
* MiSeq system is small scale sequencing system useful for amplicon sequencing and QC sequencing for large pool of RNAseq libraries form difficult material (FFPE). It can produce variety of reads length and scale dependent on type of used flowcell. It uses random clustering technology and 4 color chemistry therefore it is resistant for low complexity and low-quality sequencing libraries.
* GridIon from Oxford Nanopore system is state-of-the-art single molecule sequencing technology. It is using nanopores as an electric current detector for sequencing single stranded DNA or cDNA. It is used in our laboratory for Whole Genome Sequencing of small and medium size of genomes since can produce around 10Gb of large sequencing reads (more than 100Kb). Recently we started experiments with low and high input RNAseq on samples prepared from high quality RNA (fresh/frozen RIN >8). Using nanopore technology we are able to produce sequencing data from around 5 million full length transcripts which opens new opportunity in reproducible analysis of transcription level with detection of real splicing variants present in the sample. This application can be extended to single cell level and expanded by transition to PromethION larger scale production system.
* Chromium system from 10x Genomics is the most popular system for single cell applications. It rapidly and efficiently combines large partition numbers with a massively diverse barcode library to generate >100,000 barcode containing partitions in a matter of minutes. Therefore, system is capable to perform reproducibly single cell barcoding for many samples per day. We are using it for genome fazing, single cell ATACseq, V(D)J and 3’ end expression profiling.
* Saphyr system from BionanoGenomics is state-of-the-art optical mapping system for detection structural variation in the genomes. It can also assist in genome assembly for de novo projects. Currently we are not using this system for any type of RNAseq applications, but it can be used for verification gene fusions and other abnormalities detected in RNAseq experiments.

All sequencers are associated with on-board computers for real-time data processing, with one server dedicated for data analysis.

* To increase efficiency and reproducibility of library preparation and pooling we are utilizing six robotic systems supporting unique and overlapping applications—two Genesis systems 150 and 200 (TECAN), one Biomek (Becman Coulter), one Bravo (Agilent), one Sciclone G3 (Perkin Elmer) and Mantis (Formulatrix). Library preparation process for RNAseq application can be performed on more than one system to assure redundancy in our workflow. Additionally, we have dedicated PhD level employee extensively trained by both Tecan and Perkin Elmer to install new protocols, test them, prepare new scripts or adjust existing scripts according to current needs. We are collaborating with Tecan on development scripts for library preparation protocols on their Genesis NGS Workstation series. Our scripts for TruSeq RNA applications are released and distributed by Tecan. Here is short description of our automation systems:
* Genesis 150 (8 tip) and 200 (8 tip and 96 head) air displacement systems which are dedicated for everyday liquid handling solutions. We prepared scripts for loading 96 well plates, handling dilutions, Pico Green DNA assay and all other custom applications. Additionally, if needed, our Genesis systems are capable for TruSeq (mRNA, total RNA and RNA exome – Illumina) and NEBNext Ultra II (NEB) Stranded mRNA and total RNA library preparation up to 48 samples per batch. New library preparation protocols can be easily automated on these systems.
* Bravo from Agilent is our dedicated system only for TruSeq mRNA and TruSeq total RNA seq library preparation. We already used it for more than 15000 samples including TCGA and other large RNAseq projects. This system is capable of preparation around 384 libraries per month in full production scale. If necessary, as a redundancy for workflow, we can support production of TruSeq RNAseq libraries using both Tecan and Beckman Coulter liquid handling systems.
* Biomek (Beckman Coulter) is supporting system for our TruSeq RNAseq production on Bravo instrument. It is capable for customized liquid handling functions necessary for large production. This system is also able to prepare KAPA Stranded mRNAseq and total RNAseq libraries per our customers request.
* Siclone G3 from Parkin Elmer is high capacity 96 head system design for large volume library preparation. It is our main system for NEXT flex Small RNA library production (miRNA). We are also utilizing this system for KAPA DNA/RNA and Lexogen RNA library preparation. We can produce around 800 libraries per month in our standard production capacity. Since this system has only 96 head it is dedicated only for project over 48 samples per batch. It is also our redundancy system for TruSeq library preparation if Bravo and Beckman system are during the service.
* Mantis is miniature system dedicated for small volume liquid handling. It allows us reproducibly handle volume 0.5ul up to 20ul for selected number of processed samples at 96 plate and high speed of pipetting. Therefore, we are able to construct KAPA Hyper DNA and KAPA Hyper mRNA (during validation) libraries using portions of the reagents what reduce cost of our operation. Our R&D is planning to use this system for single cell, amplicon and RNAseq applications.

Twenty-two full-time personnel, in addition to the Managing Director, Dr. Piotr Mieczkowski, staff the HTSF. UNC Center for Bioinformatics (see below) staff is responsible for data processing, data management, initial analysis and distribution of all data generated at UNC HTSF. We have accumulated considerable experience with analysis of data from Illumina sequencers since 2007. Through collaborations with the UNC labs, RAM-Lab core and G-PATH core, we also have access to a Rhapsody (BD), additional robotics, and two NextSeq 500s.

Computer Resources

Information Technology Services (ITS) is the campus provider of leading-edge information technology service and support. The ITS division of Research Computing develops and maintains computing infrastructure for research support, while also directly engaging researchers to develop and deploy the needed tools and capabilities. Among the computing resources available for researchers at UNC-Chapel Hill are a 64-processor SGI Origin 3800 server; a 24-processor statistical computing domain in a Sun Fire 15K; a 352-processor Beowulf Linux cluster; a 32-processor SGI Origin 2400 server; and a 32-processor IBM Regatta P690 server. ITS operates several other large-scale computer systems, which are available for data management, statistical computing, and communications. An extensive library of centrally provided and managed software applications are available for use. More than two hundred software packages and utilities are offered for use on the central systems. By relying on ITS to maintain the hardware, security environment and software builds of computing systems, researchers are free to devote their time to science and research rather than to system administration. The faculty director of the HTSF, has secured the purchase an additional 64, 8 processor nodes and ~900 Tbyte of hard disk space to be dedicated to the processing and analysis of NextGen sequence data.

The UNC Center for Bioinformatics has provided Information technology and bioinformatics analysis support for the UNC-HTSF since its inception. Center for Bioinformatics staff are responsible for data processing, data management, initial analysis and distribution of all data generated at UNC HTSF. They have worked with a number of open-source software tools (including but not limited to SAMTOOLS, Bowtie, TOPHAT, BWA, SOAP2, MAQ etc.) that are used for analysis of deep-sequencing data. Hardware support for the facility includes four Dell Poweredge 2950 servers running Redhat enterprise Linux (EL4) (each with 2 x 2.66 GHz Quad core Xeon CPU’s, 32 GB RAM, 6 TB of disk storage) available for management/analysis of data from the next-gen sequencers within the Center, and a 30TB iSCSI SAN for next-gen sequence data analysis and storage. A small Linux cluster (20 node) was recently installed shortly to help facilitate large-scale analysis of sequence data. In addition to open source software programs meant for analysis of next-gen sequence data (maq, ChIP-seq, mosaik, phred/phrap/consed) there are two network licenses for CLC Genomics Workbench. Dr. Tristan De Buysscher, a senior bioinformatics scientist in the Center, provides dedicated programming support for informatics for high throughput sequencing projects.

*Updated: 4/24/2019*