Genome Analysis & Technology Core

Alyson J. Prorock[1] MS., Yongde Bao[1, 2] PhD., and Katia Sol-Church[1, 3] PhD.

[1]Genome Analysis & Technology Core, [2]Department of Microbiology, [3]Department of Pathology, School of Medicine
University of Virginia, Charlottesville VA

https://med.virginia.edu/gatc/

Introducing: The Genome Analysis & Technology Core (GATC)

The Genome Analysis and Technology Core's continuous mission is to enable high throughput Next Generation Sequencing projects for both novice and experienced users alike. The GATC serves as a scientific resource and biotechnology hub enhancing the scope and quality of basic and translational research at UVA by maintaining customer satisfaction, increasing staff accountability, collaboration and quality control. The core staff endeavor to provide timely assistance in all facets of experimentation from project design/execution to data analysis/reporting, stimulate institutional collaborations as well as provide staff/students training and educational opportunities.

The core is a fee-for-service operation that offers instrumentation and expertise in all areas of NGS genomics and transcriptomics, as well as training to access the core shared instrumentation. Current applications supported by the core staff include bulk and Single Cell (Sc) RNA Seq, CHIP Seq, ATAC Seq, Amplicon DNA Seq, 16S ribosomal gene sequencing, shot gun sequencing for small genomes and whole exome sequencing. The core also provides real time and droplet digital PCR services most suitable for targeted gene expression, SNP genotyping and CNV discovery. The GATC provides free project consultations and support for grant applications.

By opening the doors to new technologies, the core affords investigators at UVA and other partner institutions the tools for a deeper understanding of biological processes. We invite UVA students and researchers to visit us on line (google: UVA GATC) or better yet, to come and visit the labs located in Pinn Hall near the Graduate Programs Offices.

List of GATC Services

Scientific Support
- Project Design Consultation
- Grant Budget Preparation
- Grant Letter of Support

NGS Library Preparation
- 16S and shot gun metagenomics
- mRNA and total RNA library prep
- Low input RNA amplification
- Nextera XT and Nextera Flex
- Small RNA libraries

Real Time qPCR
- NGS Library quantitation
- Gene expression/validation
- SNP genotyping

Shared instrumentation Access
- Digital Droplet PCR
- Pippin
- Covaris
- QIAcube

Next Generation Sequencing
- MiSeq and Next Seq run Set up
- Library pooling services
- Sequencing runs single and paired end

10X Genomics

Single cell 3'RNA seq

Under development:
- 10X Linked read DNA
- 10X Single cell 5 RNA seq
- VDJ RNA seq

Quantitation, Preparative and QC
- NanoVue and Qubit
- BioAnalyzer
- TapeStation

Other Services
- Training and Education
- Data Analysis
- Custom libraries
- New Protocol Development

Single Cell Capabilities with 10X Genomics

The GATC's newest instrument is the 10X Genomics Controller located in Pinn Hall 1076

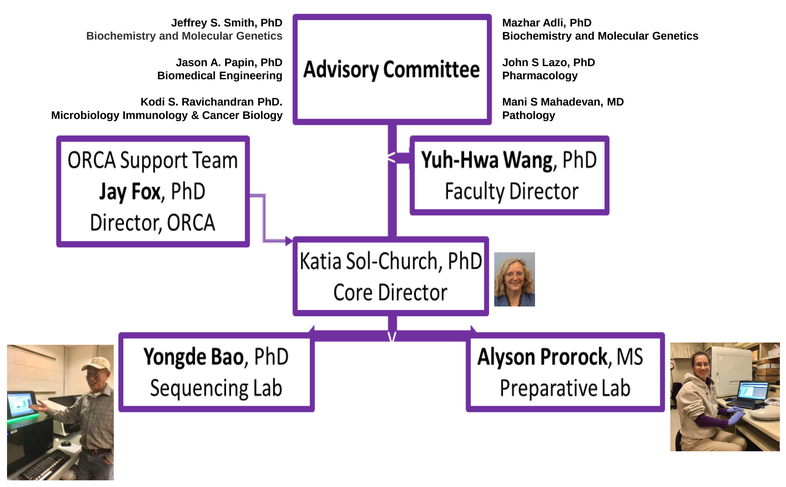
10XGenomics Chromium Controller.

Single-use microfluidics chip

The GemCode Technology's flexible throughput enables encapsulation of 500 – 80,000+ cells (8 wells) in 10 minutes and supports processing of (ready to sequence) barcoded cDNA libraries in two days

- This technology is compatible with the core's Illumina NextSeq/MiSeq sequencers and can be used as standalone for library prep or integrated to GATC in house sequencing workflow.
- At this time the GATC core enables RNA expression profiling of hundreds to thousands single cells. A typical experimental set up will analyzed the expression of 1,000 to 2,000 cell at a depth of 50,000 to 100,000 reads per sample.
- Kits are currently available for haplotype analysis in HMW DNA using barcoded linked reads.
- New protocols are being developed to enable VDJ repertoire typing of B and T cells paralleled with 5' RNA expression profiling.

GATC Organizational Chart



How to Match Your NGS Research Goal with Core Technologies

RNA sequencing

RNA-Seq experiments should be performed with at least two or more biological replicates. The first step in any successful sequencing experiment is the preparation of the RNA to be sequenced. The number of RNA samples that can be analyzed on the core's sequencers will depend on RNA quality, depth of sequencing needed (Goal) and Output of the sequencing kit. Next Seq 500: Mid Output (MO) Kit generates 130 million reads, and High Output (HO) Kit 400 million reads. Mi-Seq V2 and V3 sequencing kits can generate respectively 17 and 25 million reads per run. Below are recommendations for typical RNA-Seq applications.

Goal #1 : I want to focus on the coding transcriptome and I want to quantify gene expression at the gene level, with one abundance value generated per gene.	**Method: Gene expression Profiling – mRNA-seq** **≥ 20 Million reads per sample, 1 x 75 bp** **Library prep: mRNA stranded** **Next Seq MO mode: 7 sample pool** **Next Seq HO mode: 20 sample pool** **MiSeq: 2 samples can be pooled**
Goal #2 : I want to focus on the RNA exome and I want to quantify gene expression by analyzing abundance values for every transcript isoform from each and identify novel transcript isoforms, SNVs, gene fusions, and/or identify allele-specific expression.	**Method: mRNA-Seq** **≥ 25 Millions reads per sample, 2 x 75 bp** **Library prep: mRNA stranded** **Next Seq MO mode: 5 sample pool** **Next Seq HO mode: 16 sample pool** **MiSeq: 1 sample per flow cell per run**
Goal #3: I want to focus on the abundance values of both coding and multiple forms of noncoding RNA and identify novel transcript isoforms, SNVs, gene fusions, and/or identify allele-specific expression.	**Method: Total RNA Sequencing – rRNA depleted** **≥ 50 Million reads for QC samples, 2 x 75 bp** **≥ 100 Million for degraded samples, 2 x 75 bp** **Library prep: stranded total RNA ribo-depletion** **Next Seq MO mode: 2 sample pool** **Next Seq HO mode: 8 sample pool**

DNA Sequencing

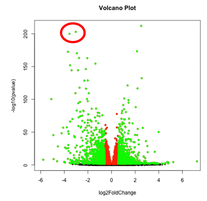
How to estimate and achieve the desired NGS Coverage for DNA sequencing will depends on the application used and best practice as recommended by the scientific community. Whole genome recommendation is 10X to 30X, while CHIP-Seq is 100X.The Lander/Waterman equation is a method for computing coverage: C = LN / G. Thus, the total number of reads needed N =CXG:L, where C is the coverage, G size of haploid genome and L is the read length (e.g.75 base-long reads). Examples of popular DNA sequencing applications are shown below (inquire for additional examples).

Goal #1: I want to sequence the known exome at 50X mean coverage	**Method: Whole Exome Sequencing WES** **Next Seq MO mode: 3 sample** **Next Seq HO mode: 12 sample pool**
Goal #2: I want to sequence a small genome 130 Mb at 30X coverage 2X150bp	**Method: Small Whole Genome Sequencing** **Next Seq MO mode: 10 sample** **Next Seq HO mode: 30 sample pool**
Goal #3: I want to sequence 12Mb region at 20X	**Method: targeted sequencing** **Next Seq MO mode: 12 sample** **Next Seq HO mode: 36 sample pool**
Goal #4: I want to sequence the 16S bacterial ribosomal genes	**Methods: Metagenomics 16S** **MiSeq 600 Cycle kit, 25 Million reads** **Up to 96 samples**

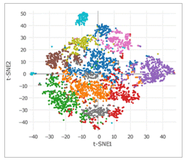
See What The Core Can Do For You

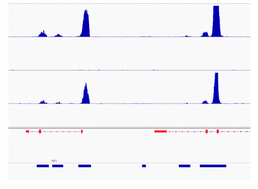
The GATC data analysis services include Differential Gene Expression (Panel A), Single Cell 3' Gene Expression (Panel B), CHIP-Seq Analysis (Panel C), 16S Metagenomics (Panel D).

Panel A: RNA-Seq
Volcano plot depicting gene expression resulting from a cell perturbation. Examples of genes that display both large magnitude fold-changes and high statistical significance p-value are circled.

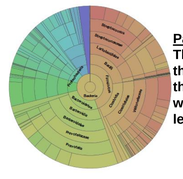


Panel B: Gene Expression
Visualization of High Dimensional Single Cell RNA-seq: the T-SNE plot identifies discrete subpopulations of differentially expressed Neurons.





Panel C: CHIP-Seq
Examples of the profiles generated by chromatin immunoprecipitation followed by sequencing (ChIP–seq). The genome browser snapshot shows the binding profiles of the factor within the region of interest.



Panel D: 16S typing
This sunburst chart shows the relative abundance of the classification results within each taxonomic level.

Next Generation Sequencing Instrumentation

The Genome Analysis and Technology Core NGS services located in UVA's Pinn Hall 1044 include two benchtop Illumina sequencers.

Illumina MiSeq Sequencing System

The Miseq enables automated paired-end reads and up to 15 Gb per run, delivering over 600 bases of sequence data per read.

The core provides optimized library prep kits for a variety of applications
- Targeted resequencing and CHIP-Seq
- 16S/metagenomics and small genome sequencing
- Targeted gene expression profiling

Illumina NextSeq™ 500 Sequencing System

Its flexible configuration (Mid and High output) enables 120Gb of output with 130-400 Million reads per run.
- Shotgun/de novo sequencing
- RNAseq bulk or single cell
- Exome sequencing
- Small RNA profiling as well as other custom NGS pipelines

GATC Technology Hub

Agilent Tape Station

DNA and RNA QC

Qubit Fluorometer

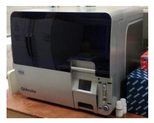
NanoVue Spectrometer

BioRad QX200 ddPCR

Copy number Library quant

Agilent Bioanalyzer

Real Time PCR System



Qiacube Sample prep

Pippin Size Select

Covaris DNA Shearing