Announcements:

Join us for the GSAF Spring Lunch and Learn Series

Single-cell RNA Sequencing (Illumina)
- Wednesday April 12, 11am - 12:30pm
- MBB 1.210
- Next-Generation Sequencing Tools for Single-Cell Gene Expression and Transcriptome Analysis
- Mehdi Keddache, PhD, Sequencing and Data Analysis Specialist, Illumina Inc.
- More information at Illumina's Event Registration Page

Long-read Sequencing Technology and Applications (Oxford Nanopore Technologies)
- Tuesday April 25, 12 noon - 1:30pm
- NHB 1.720
- Real time DNA sequencing using Oxford Nanopore Technologies ‘Nanopore Sensing’ Platform
- James Brayer, Associate Director, Market Development

Oxford Nanopore Technologies has developed a disruptive platform for the direct, electronic analysis of single molecules. Our instruments the MinIONTM, the PromethIONTM and the GridION X5 ™ are adaptable for the detection and analysis of a range of analytes that include DNA, RNA, proteins and small molecules. At the heart of our platform is a biological protein called a ‘nanopore’. A single nanopore create a hole in a membrane made from a proprietary synthetic polymer. An electric potential is applied across the membrane resulting in a current flowing only through the aperture of the nanopore. Single molecules that enter the nanopore cause characteristic disruptions in the current, by measuring these disruptions single molecules from a sample are identified. The MinION is a small device that is designed for portability and simplicity of its workflow. The MinION plugs into a standard PC or laptop using the USB port. GridION X5 is a compact benchtop system designed to run and analyse up to five MinION Flow Cells. The PromethION is a standalone high throughput benchtop instrument that provides the flexibility to run up to 192 libraries in an asynchronous manner. This allows for large projects that requires the flexibility and throughput to interrogate complex eukaryotic genomes. Oxford Nanopore is integrating the data produced by the MinION, the GridION X5 and the PromethION into a cloud-based analytics company, Metrichor. Metrichor is powered by its EPI2ME platform. Metrichor is providing tools to automate data analysis workflows to help people track, trend and predict biological data resulting in real time actionable interpretation of their data. Users of the technology have access to our ‘Nanopore Community’. The Nanopore Community helps new users get started with technical documentation as well as user driven forums and encourages discussion and collaborative experimentation using our technology. There is a growing list of publications on the many uses for our nanopore sensing platform that include field based applications, real time pathogen detection and surveillance, metagenomics analysis, anti-microbial resistance detection, education and many more including sequencing on the International Space Station. I look forward to sharing with you the unique opportunities enabled by our nanopore sensing approach.

- Reserve your lunch at https://utexas.qualtrics.com/jfe/form/SV_0O1FrR6CsHkL941

Long-read Sequencing Technology and Applications (Pacific Biosciences)
- Wednesday April 26, 11am - 1:00 pm
Index Switching Preprint Concerns Sequencing Community *Update-19Apr17 See Below for Notes from Illumina

What is Index Switching?

A preprint released April 9, 2017 by Sinha, et. al., describes a “spreading-of-signal” phenomenon that is attributed with causing 5-10% of sequencing reads to be incorrectly assigned in situations where multiplexed libraries are run on Illumina’s HiSeq 3000/4000 systems. The observed errors are restricted to the HiSeq 3000/4000 and X Ten systems that use the exclusion amplification method for generating clusters in the nanowells of patterned flow cells. According to the pre-print titled: “Index switching causes ‘spreading-of-signal’ among multiplexed samples in Illumina HiSeq 4000 DNA sequencing,” low levels of free index primers in the pool get extended by DNA polymerase to create a new library molecules in the Cbot during the initial stage of cluster generation but prior to binding to the patterned flow cell.

What is the extent of the problem?

According to the authors, the RNA-seq experiments in this study revealed a 5-10% error rate associated with “signal spreading.” In a commentar y in the Molecular Ecologist, Ethan Linck, proposed that the problem is most likely to be troublesome for multiplexed RNA-seq studies or “studies attempting extremely low frequency variant detection, where a handful of erroneously indexed reads could have a big impact on inferences.” Studies not affected by the phenomenon include non-multiplexed studies, and dual-indexed samples where each end is unique. Clearly, more data is needed under different situations and study designs to gain a better understanding of the extent of this potential problem.

What is being done?

Soon after the pre-print was released on April 9th, Illumina responded by tweeting, “we’re aware & working on it. Data indicates it occurs at low rates,and impact may be mitigated w/other index approaches. Correcting this is a high priority and we are evaluating fixes. Pls keep feedback coming via DM, your Illumina team or email (2/2).” The GSAF wants to reassure our customers that we are actively looking into this issue and will be working on alternatives to our current indexing system with the continual goal of delivering highly accurate sequencing data.

White Paper and Best Practices Release from Illumina

Best Practices from Illumina

Effects of Index Misassignment on Multiplexing and Downstream Analysis

Getting started

New to the GSAF? Expand here to get started...

1. Review this self-paced training presentation.
2. Read this page further to learn what instruments we have and what we’re experienced with.
3. Sign-up on our mailing list by joining "gsafusers" at UT's list server - just select “subscribe” and enter "gsafusers".
4. If you want to submit samples for the first time, first get a UT EID if you don’t have one, then go to our sample submission section for more info, to check our queues, etc..

Get directions here
Protocols, instruments, computers, and software of the UT GSAF

The GSAF serves a wide array of customers and has gained considerable experience in preparing samples and analyzing data for NGS projects. We have industry-leading platforms and some user-accessible instruments. If you're just considering a project or are ready to sequence, we're here to give you great data quickly and affordably. Click to expand...

The GSAF is housed in approximately 2,000 square feet of controlled-access laboratory space and is an experienced NGS facility. Launched in 2008, the GSAF currently processes over 5,000 NGS samples per year.

Our equipment and capabilities include:

1. Two Illumina HiSeq 2500 next-generation DNA sequencers and associated peripherals. The HiSeq can generate over 600 gigabases of sequence in an 11-day run or 120 gigabases in a ~1 day rapid run. The GSAF has experience generating small RNA, mRNA, genomic DNA fragment, RAD (including ddRAD), bacterial and fungal metagenomics, and genomic DNA large-insert mate-pair libraries for the Illumina platform.
2. Two Illumina MiSeq next-generation DNA sequencers and associated peripherals. The MiSeq is intended for lower data output, faster turn-around time projects, or for projects requiring longer read lengths (up to 600 bp per template, as two 300 bp sequences).
3. Informatic tools and hardware sufficient for analysis of next-generation DNA sequencing data, including:
   a. Multiple Dell PowerEdge R900 & R720 servers with 16-32 cores each, with BWA, Bowtie, Shrimp, SOAP, and mapreads aligners, NCBI BLAST, and R/Bioconductor, and several de novo assemblers, sequence editors, and genome browsers.
   b. The resources of Lonestar, a 22,656 core supercomputer, and Stampede, a 400,000 core supercomputer, both hosted by UT’s Texas Advanced Computing Center, with 15 petabyte of scratch disk storage.
   c. The GSAF has software stacks and experience appropriate to your project and both we and TACC staff can assist you in training your students to use this resource effectively.

Lab protocols we are experienced with:

- Creation of fragment (single-end or paired-end) sequencing libraries for all Illumina next-gen sequencers
- Creation of mate-pair or "jumping" sequencing libraries between 1.5kb and 4kb for Illumina next-gen sequencers
- Creation of RNA-seq libraries from total RNA, small RNA, and immuno-precipitated RNA for Illumina next-gen sequencers
- Creation of amplicon libraries (particularly 16s and ITS regions) for the Illumina next-gen sequencers
- Human exome and custom capture with the Agilent SureSelect, Illumina TruSeq, and Nimblegen SeqCap EZ kits
- Sample and library QC using the Agilent BioAnalyzer, Picogreen and Ribogreen fluorimetry, qPCR, and spectrophotometry

Bioinformatic protocols we are experienced with:

- RNA-seq for transcript abundance, alternative splicing analysis, and variant detection
- SNP/variant analysis
- small RNA abundance and alternative editing analysis
- de novo and reference-guided assembly from fragment, paired-end, and mate-pair data on both DNA and RNA (transcriptomes)
- Whole exome data analysis

Instruments in our lab:

- Illumina HiSeq 2500 sequencers (two)
- Illumina MiSeq sequencers (two)
- Illumina NextSeq 500 sequencer (one)
- Covaris S220 Adaptive Focused Acoustic shearing device
- DigiLab HydroShear shearing device
- Agilent BioAnalyzer 2100
- Agilent TapeStation
- Invitrogen Qubit fluorimeter

Computational and software resources:

- The GSAF hosts a Dell R720 32-core, 196 GB server with a total of 74 TB local disk dedicated to NGS analysis. Access is available free of charge to all GSAF customers. We maintain a wide range of tools for NGS analysis and assembly on this server. Here are instructions to Getting an account on our server.
- In addition, the GSAF uses and works with the TACC bioinformatics group, supporting tools and applications suitable to the TACC environment.
- Want to get started? Contact us if we can help, or here's some documentation describing how to submit samples to the GSAF.
- Need to start analyzing your data? Get an account on our computational server. New to Unix? Check out some Unix and Perl resources for beginners.

Need Accounts? Here's how you can get an account on the GSAF server

Welcome to the UT GSAF Wiki - a central source of information for next-gen sequencing at UT Austin.

All are welcome to explore. With a UT EID you may also edit and contribute. Please sign up for our user group email to get notifications of upcoming seminars, NGS club meetings, and changes to service. Just hit "subscribe", enter your email address, and "submit". You don't have to
Reference Pages (GSAF and user contributed)

Browse the wiki:

**Software and Reference Genomes**

The [Software](#) section of the BioITeam wiki site lists software available on the GSAF Server Fourierseq and/or TACC and how to use it. The [Reference Genomes and Databases](#) section lists pre-installed reference genomes and their mapping indexes.

**Lab Protocols and Oligo Sequences**

[Lab Protocols](#) and [Oligo Sequences](#) are useful if you are designing experiments or preparing your own libraries.

**User Project Analysis Pages**

Use [User Project Pages](#) to work with the GSAF about an ongoing analysis project - [user editable](#), share with your PI! Note that in-process laboratory status updates are not stored here.

**How to...**

- [How to submit samples to the GSAF](#)
- [How to download your data](#)

Thanks to [Craig Dupree](#) of CCBB for pointing us to the UT Wikis infrastructure!