Announcements:

4April18: GSAF introduces a new library prep service, TagSeq based on the published methods below. TagSeq is a 3' RNA library prep method, please visit our pricing page for more details.

- Evaluation of TagSeq, a reliable lowcost alternative for RNaseq
- Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNaseq procedure

19April17 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...**

  - Review this self-paced training presentation.
  - Read this page further to learn what instruments we have and what we're experienced with.
  - Sign-up on our mailing list by joining "gsafusers" at UT's list server - just select "subscribe" and enter "gsafusers".
  - 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:

  - 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.

  - 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).

  Informatic tools and hardware sufficient for analysis of next-generation DNA sequencing data, including:

  1. 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.

  2. 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.

  3. 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 be at UT to subscribe.

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!