THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

# Laboratory-based Next Generation sequencing didactics in an undergraduate environment


**Vibhuti Srivastava Ph.D., Peter Hu, Ph.D., Irene Newsham Ph.D.**

**Molecular Genetic Technology Program, School of Health Professions, MD Anderson Cancer Center**


## Introduction

The Molecular Genetic Technology program in MD Anderson's School of Health Professions is accredited by National Accreditation Agency for Clinical Laboratory Sciences and awards a Bachelor of Science degree. This program integrates theoretical, conceptual and practical aspects of key molecular diagnostic techniques. Next generation sequencing (NGS) is one such technique that is increasingly being utilized for clinical testing. Although it has a broad range of clinical applications, the ability of clinical and allied health undergraduate programs to provide hands-on experience with NGS is limited by funding, availability of trained instructors and a lack of a practical laboratory-based training module.

## Objectives

1. Develop a cost effective NGS laboratory training module suitable for implementation in an undergraduate environment
2. Provide undergraduates in the MGT program at MDACC a hands-on training experience on the Ion Torrent (PGM) NexGen sequencing platform.
3. Generate NGS data on clinical samples and perform preliminary analyses using publicly available software suite
4. Identify the challenges and limitations of NGS module implementation under the constraints of an undergraduate program

## Methods

1. We first developed a hands-on Ion Torrent NGS module as a part of an advanced laboratory techniques course in 2014.
2. During a two-week long module, each student prepares a bar-coded library from a cancer sample procured through our clinical collaborator using the Ion AmpliSeq™ Cancer hotspot panel.
3. Students perform all the experimental steps individually until emulsion PCR.
4. A group of two to four students pool their barcoded libraries and perform subsequent steps of clonal amplification, target enrichment and sequencing.
5. The generated sequencing data from Ion reporter software is discussed and preliminary data analyses is performed using GALAXY, a web based open source platform, as a part of a concurrent Bioinformatics course.
6. Assessment of learning outcomes is based on an associated lab report.

### *Ion Torrent Laboratory Practical-Scheduling*

PGM Library Prep
Each Team will make ONE PGM barcoded library.

Emulsion PCR/Sequencing
Two or three libraries will be combined to perform the emulsion PCR (OT2), enrichment and sequencing runs

Sequencing Run Groups:
Group A: Teams 1-3
Group B: Teams 4-5
Group C: Teams 6-8
Group D: Teams 9-10

Please Note:
All times are approximate. You are expected to stay for the time required to successfully perform the sequencing lab, unless you have previously informed the faculty about a prior issue or commitment.

| Monday | Tuesday | Wednesday | Thursday | Friday |
|---|---|---|---|---|
| | Library Prep - all | Library Prep - all | | |
| 11 | 12 OT2 Emul PCR: 8-9 pm Group A<br>OT2 Emul PCR: 4pm-OIN Group B | 13 NOT AVAILABLE | 14 | 15 OT2 Emul PCR: 8-9pm Group D<br>Enrichment: 9am-10:30 am; Group C<br>Enrichment: 10:30am-12pm; Group B<br>Enrichment: 12pm-1:30 pm; Group A<br>Enrichment: 3:30pm-5pm; Group D<br>OT2 Emul PCR: 3 pm - OIN Group C |
| 18 Clean/Initialize PGM: 8-9am IN/VS<br>Sequencing: 9-1pm Group A | 19 Clean/Initialize PGM: 8-9am IN/VS<br>Sequencing: 9-1pm Group D | 20 Sequencing: 9-1pm Group B<br>Sequencing: 1-5pm Group C<br>NOT AVAILABLE | 21 PGM Analysis Lab VS 1 – 5pm | 22 NOT AVAILABLE |

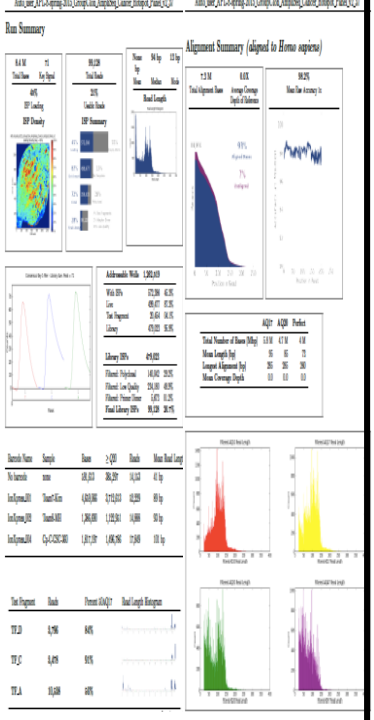
### *AmpliSeq™ Cancer panel details*

- Requires only 10 ng of FFPE or higher-quality gDNA, yields library in ~3.5 hours
- Panel targets >2,800 COSMIC mutations in 50 cancer-associated genes
- A single tube of primers for 207 amplicons (avg length= 154 bp)
- Samples can be barcoded and library-prep automated for multiplexing

**The 50 targeted genes**

| | | | |
|---|---|---|---|
| ABL1 | EZH2 | JAK3 | PTEN |
| AKT1 | FBXW7 | IDH2 | PTPN11 |
| ALK | FGFR1 | KDR | RB1 |
| APC | FGFR2 | KIT | RET |
| ATM | FGFR3 | KRAS | SMAD4 |
| BRAF | FLT3 | MET | SMARCB1 |
| CDH1 | GNA11 | MLH1 | SMO |
| CDKN2A | GNAS | MPL | SRC |
| CSF1R | GNAQ | NOTCH1 | STK11 |
| CTNNB1 | HNF1A | NPM1 | TP53 |
| EGFR | HRAS | NRAS | VHL |
| ERBB2 | IDH1 | PDGFRA | |
| ERBB4 | JAK2 | PIK3CA | |

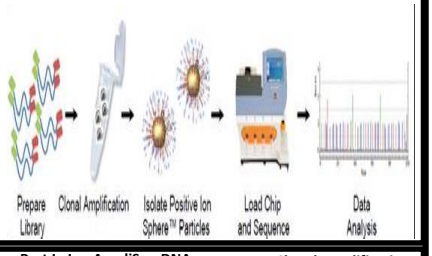
### *Sample Sequencing Run Data and Preliminary Analysis*




### *Preview of low frequency variant calls for one GBM sample*

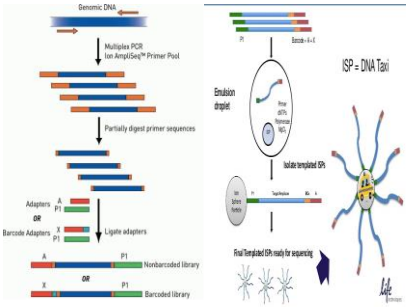
| Chrom | Position | Ref | Variant | Allele Call | Filter | Frequency | Quality | Filter | Type | Allele Source | Allele Name | Gene ID |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chr3 | 178947827 | G | T | Heterozygous | | 46 | 458.98 | - | SNP | Novel | -- | N/A |
| chr4 | 55142355 | A | G | Homozygous | | 100 | 542.27 | - | SNP | Novel | -- | N/A |
| chr4 | 55972974 | T | A | Heterozygous | | 59.5 | 517.28 | - | SNP | Novel | -- | N/A |
| chr5 | 112179770 | G | A | Heterozygous | | 96.4 | 1254.82 | - | SNP | Novel | -- | N/A |
| chr7 | 55249063 | G | A | Heterozygous | | 45.2 | 182.99 | - | SNP | Novel | -- | N/A |
| chr9 | 33675753 | A | G | Homozygous | | 100 | 272.55 | - | SNP | Novel | -- | N/A |
| chr10 | 43812843 | G | T | Homozygous | | 100 | 349.74 | - | SNP | Novel | -- | N/A |
| chr13 | 28612282 | T | C | Heterozygous | | 54.3 | 377.91 | - | SNP | Novel | -- | N/A |
| chr13 | 28612183 | A | G | Homozygous | | 100 | 1116.97 | - | SNP | Novel | -- | N/A |
| chr17 | 7577548 | C | T | Homozygous | | 100 | 990.65 | - | SNP | Novel | -- | N/A |
| chr22 | 17715458 | G | A | Heterozygous | | 36.7 | 85.27 | - | SNP | Novel | -- | N/A |

### Ion Torrent Laboratory Practical - Workflow

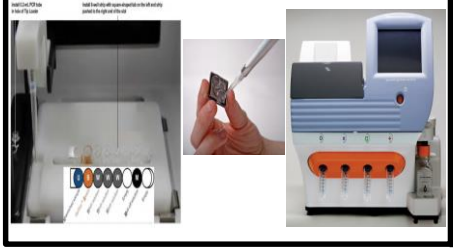



### Part I: Ion AmpliSeq DNA Barcoded Library

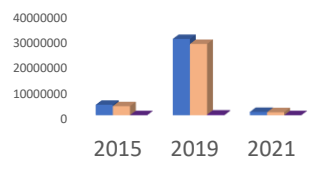
### Part II: Clonal Amplification by Emulsion PCR



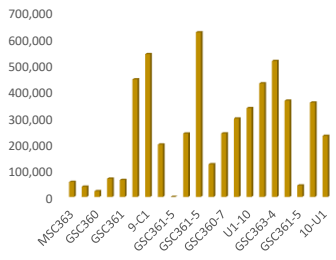

### Part III: Library Enrichment and Sequencing



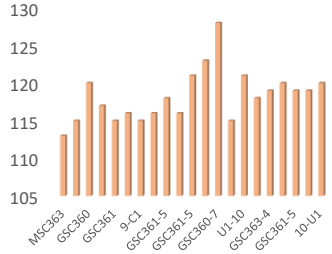
### Year*-wise data comparison




**Reads per sample- 2021**




**Mean Read Length(bp)-2021**




* NGS could not be performed in 2016,2017 due to extreme weather, 2020: COVID

## Results

1. A total of 87 students participated, generating 57 libraries in 2015, 2018-2021.
2. Overall, 10/57(17.5%) libraries failed with zero sequenced bases.
3. Data quality was variable over individual academic years.
4. Highest average number of bases with a ≥Q20 score obtained in 2019 and lowest in 2018.

## Successes

1. Students gained significant conceptual understanding of the logistics of the NGS process in this laboratory practical as measured on associated lab reports and exams.
2. No appreciable difference in measured learning was found whether library construction occurred in teams or individually.
3. Graduating students have been hired onto sequencing teams based on their lab experience in the MGT program
4. Sequencing with expired reagents produced sufficient data for downstream bioinformatic analysis.

## Challenges

1. Undergraduate student's schedule and laboratory availability and time constraints
2. Complexity of the protocols and student bench skills
3. Quality of run data inconsistent and not high enough quality to publish
4. Chip loading/Sequencing run costs and reagent expiration

## Conclusion

We demonstrated the viability and sustainability of a cost effective NGS experiment in a didactic setting as a part of an undergraduate diagnostic science curriculum. Variability of data quality was expected as students had no prior experience with this technique and is in part due to the age of available sequencing reagents. Introduction of a hands-on NexGen sequencing experience at the undergraduate level is expected to prepare molecular genetic technologists for a smooth transition to NGS-based roles in clinical molecular diagnostic labs.