DNA REPAIR DATABASE

STUDENTS / UNIVERSITIES

Rumeysa Hanife KARS/Istanbul Medipol University
Selen UYGUN/ Sabanci University
Hatice ARSLAN/Kadir Has University

SUPERVISOR(S)

Ogun ADEBALI

ABSTRACT

In DNA Database project our main goal is the detection of rare mutations and site-specific damage in DNA. Since it is now possible due to the developments in high-throughput sequencing technologies, however, the process' error rates have constructed a major obstacle and generate high noise thresholds for detection.To extinguish the error rates, a variety of modified library constructions have been developed. These helped the process by creating redundant sequences from fragments of DNA.. By using modified library constructions, our group have been trying to reduce the error rates so as to get low error-rate data and aimed to create pipelines by using these data.[1]

OBJECTIVES

- To determine damage sites identified by different sequencing methods.
- To create a publicly accessible DNA repair database and publish data obtained from this research.

SEQUENCING-BASED METHODS

1-CISPLATIN-SEQ

- Cisplatin is one of the widely used drug to cure cancer.
- Its aim is to kill cancer cells by damaging their DNA.
- Cisplatin has a central role in cancer chemotherapy. It is effective in the healing tumors and also several types of cancer.[2]

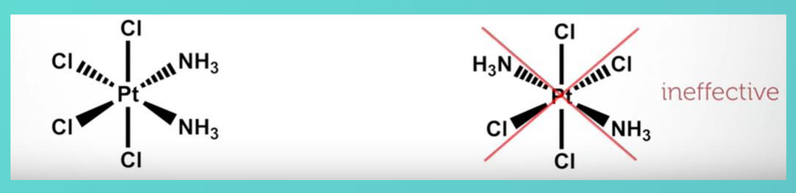

Figure 2. Cis-platin and trans-platin [3]

- Cisplatin is in the platinum-based medicines. It works by linking to DNA and preventing its replication.

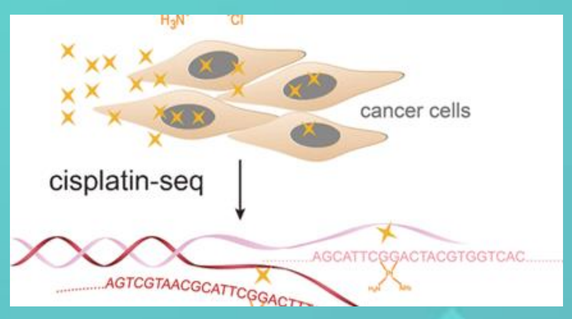

Figure 3. Cisplatin-seq process[4]

2-PBAT (Post Bisulfite Adaptor Tagging)

- DNA methylation plays a critical role in epigenetic regulation of the genome.
- Bisulfite treatment before sequencing is commonly used to determine the pattern of methylation in DNA. After bisulfite treatment, unmethylated cytosine nucleotides convert to uracil nucleotides while methylated ones remain the same.
- PBAT attaches random adaptors to bisulfite-converted genomic DNA and enables bisulfite-sequencing from sub nanogram quantities of DNA.[5]

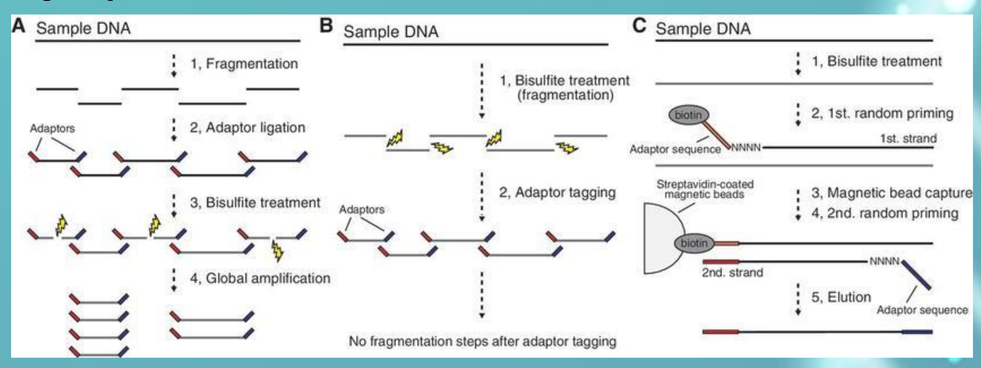

Figure 4. Common bisulfite-sequencing protocols[6]

3- RIBOSE-SEQ

Yeast is used on experiments on ribose-sequencing method and results showed us the parts of ribonucleotide incorporation in mitocondrial and nuclear DNA.The detection of ribonucleotide in Dna is significant since normally Rnase H2 prevents the incorporation however its incorporation is possible in the case of inactivation of Rnase H2. This method captures elements at the alkaline cleveage of DNA at embedded ribonucleotide.Additionally spotting these parts are important for neural disease treatments. Extra Hydroxyl groups are the keys in the technique due to the non-random distribution of them in DNA.[7]

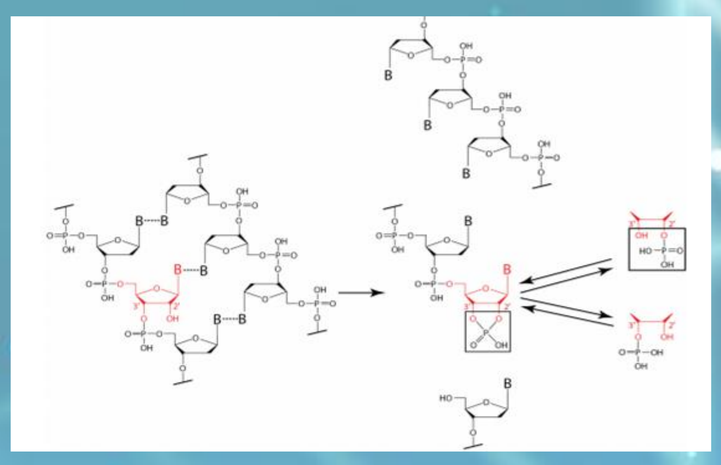

Figure 5. Mechanism of alkaline cleavage of ribonucleotides in DNA[8]

4-EMRIBOSEQ

•To determine the location of embedded ribonucleotides in the genome, emRiboSeq is established. It is based on the separation of the phosphodiester bond with RNase H2.
•RNase H2 is one of the enzymes, and it provides the eliminations of misincorporated ribonucleotides.[9]

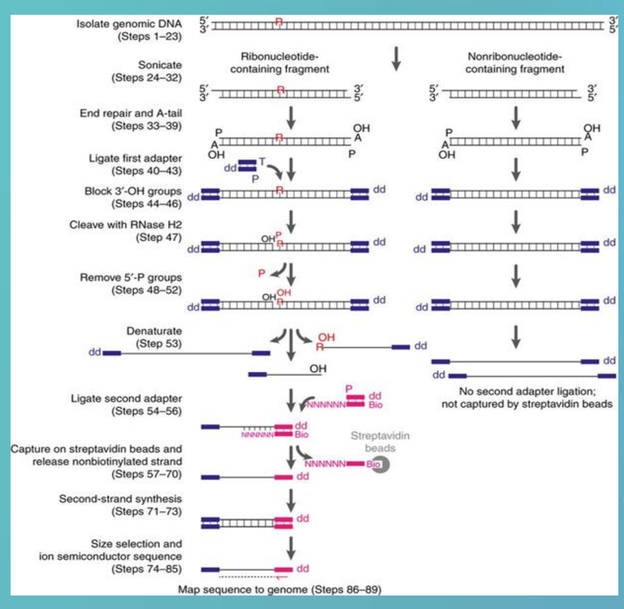

Figure 5. The process of emRiboSeq [10]

BASIC COMMANDS

.fastq / .fasta

- Downloading read and reference genome sequence files.

```
fastq-dump ${NM}.fastq
```

.fastq

- Trimming adaptors and low quality bases

```
java -jar <path to trimmomatic.jar> PE [-threads <threads] [-phred33 | -phred64] [-trimlog
 <logFile>] >] [-basein <inputBase> | <input 1> <input 2>] [-baseout <outputBase> | <unpaired
 output 1> <paired output 2> <unpaired output 2> <step 1> ...
```

.sam

- Alignment of reads and reference genome sequence

```
bowtie2 -x reference_genome -U read_1.fastq -S ${NM}.sam
```

.bam

- Converting .sam file to .bam file which is a more storable format.

```
samtools view -b -S ${NM}.sam > ${NM}.bam
```

.bed

- Converting bam to bed which is a more convenient format to run with other data files.

```
bamToBed -i ${NM}.bam > ${NM}.bed
```

CONCLUSION

Continuous improvements in this area, including the use of bioinformatics data/statistical error-correction methodologies should maintain to expand usage. Up to this point, applications of these sequencing methods have centered widely on medical and biomedical purposes, especially those relevant to cancer biology. We will continue our studies with the evaluation of our methods to create a pipeline.

REFERENCES

[1] Sloan, D., Broz, A., Sharbrough, J. and Wu, Z. (2018). Detecting Rare Mutations and DNA Damage with Sequencing-Based Methods. Trends in Biotechnology, 36(7), pp.729-740.
[2] Shu, X., Xiong, X., Song, J., He, C. and Yi, C. (2016). Base-Resolution Analysis of Cisplatin-DNA Adducts at the Genome Scale. Angewandte Chemie, 128(46), pp.14458-14461.
[3]Biolegend.com. (2019). BioLegend Blog - From Bugs to Bedside – The Rise of Cisplatin as an Anti-Cancer Agent. [online] Available at: https://www.biolegend.com/newsdetail/4339/ [Accessed 1 Aug. 2019].
[4] Phys.org. (2019). Identification of genome-wide cisplatin cross-linking sites with DNA base resolution. [online] Available at: https://phys.org/news/2016-10-identification-genome-wide-cisplatin-cross-linking-sites.html [Accessed 1 Aug. 2019].
[5][6] Miura, F., Enomoto, Y., Dairiki, R. and Ito, T. (2012). Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Research, 40(17), pp.e136-e136.
[7] [8] Koh, K., Balachander, S., Hesselberth, J. and Storici, F. (2015). Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. Nature Methods, 12(3), pp.251-257.
[9][10] Ding, J., Taylor, M., Jackson, A. and Reijns, M. (2015). Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. Nature Protocols, 10(9), pp.1433-1444.