



ADOPTING POWER BI IN IDENTIFYING DNA SEQUENCES USING FANSE3

¹Kurumaddala Veda Srija Mahalaxmi, ²Kandhati Tulasi Krishna Kumar

¹3rd Bachelor of Pharmacy, ²Associate Professor & Placement Officer

¹Srinivasa Rao College of Pharmacy, Visakhapatnam, India.

Abstract: For diagnosis and treatment of disease, it'll usually take a long period of time until unless it is diagnosed at the initial stage. With genomic data(DNA sequences), the disease can be detected very soon, even prior to the symptom are shown. To read or map the DNA sequence any method or technology is needed. Among Sanger sequencing, capillary electrophoresis, Next Generation Sequencing, DNA microarray technology (DNA chip), most used and robust application is explained. And using FANSe3 algorithm, reports are generated and produced reports are analyzed using Power BI. This will help to detect the disease at very possible initial phase.

Index Terms - DNA, RNA sequencing, genetic code, FANSe3 algorithm, Next Generation Sequencing, Power BI, traditional medical reports, FASTQ file format

I. INTRODUCTION

To identify and diagnose diseases, different comparison studies are conducted, with public health being of ultimate importance. Clinical studies are these patient-to-patient comparisons that may include information on demographics(like analytical, arithmetical and numerical), diagnosis, co-morbid conditions, symptoms, and medications. Genomic (DNA sequences) ^[23] data can be used to detect an illness immensely early, even before the symptoms appear, allowing for early diagnosis and therapy. The order of nucleotides ^[15] is determined through DNA sequencing. DNA sequences are the nucleotides, A, G, C, and T, that make up a DNA molecule. These four fundamental building blocks can vary in at least one way, causing a difference or disease. The following is a list of a few of the abnormalities.

s.no	Genetic disorder ^[18]	gene	Repeat pattern ^[8]	location	mutation	Severity	Symptom
1	Myotonic dystrophy-1	DMP K	CTG	3'UTR	expansion	Severe	Muscle stiffness, clouding of eye lens
2	Breast cancer	EGFR	CA	intron	expansion	Severe	Lumps in nipple, nipple discharge
3	Huntington disease	HTT	CAG	Coding	expansion	Mild-severe	Involuntary jerks, unusual eye movements
4	Spinocerebellar ataxia	ATXN 3	CAG	Coding	expansion	Mild-severe	Unclear speech, difficulty swallowing
4998	Fragile x syndrome	FMR1	GCG	5'UTR	Expansion	mild	Narrow face, large ears
4999	Hepatocellular	AR	CAG	Coding	Deletion	severe	Upper abdominal pain, yellowing of skin

	carcinoma						
5000	Occulopharyngeal muscular dystrophy	PABPN1	GCG	Coding	Expansion	Mild-severe	Tongue shrinking, pooling of saliva

DNA sequencing is the process of determining the order of nucleotides (A, C, G, and T) in a DNA molecule. It has had a significant effect on a variety of biological, medical, and biotechnological disciplines. The importance of DNA sequencing includes the following.

1.1 Understanding of Genetic Information: Understanding the genetic makeup of diverse creatures, including humans, has been made possible by DNA sequencing. This has upgraded our knowledge of how genes function and how genetic variants affect treatment responsiveness and illness vulnerability.

1.2 Disease Diagnosis: Genetic illnesses like cystic fibrosis, sickle cell anemia, and Huntington's disease can now be diagnosed thanks to DNA sequencing. With DNA sequencing, clinicians can pinpoint the mutations that lead to these diseases and create individualized treatment programmes for each patient.

1.3 Drug Development: By enabling researchers to find new drug targets and evaluate a huge number of molecules, DNA sequencing has sped up the drug development process.

II. Existing System

Adenine, guanine, thymine, and cytosine are the four chemical bases that are present in a DNA^[30]. Each base is joined with the help of a sugar and phosphate molecule. A nucleotide is made up of a base, a sugar, and a phosphate. The double helix is a spiral shape formed by two long strands of nucleotides. Thiamine and adenine bind together, while cytosine binds to guanine. The arrangement of these bases creates a unique genetic code, which has an impact on the body's ability to produce various chemicals and proteins. Any alteration to these sequences could cause the specific protein to breakdown. The genetic code can be illustrated as below:

	U	C	A	G	
U	UUU } Phenylalanine (Phe)	UCU } Serine (Ser)	UAU } Tyrosine (Tyr)	UGU } Cysteine (Cys)	U
	UUC }	UCC }	UAC }	UGC }	C
	UUA } Leucine (Leu)	UCA }	UAA } Stop	UGA } Stop	A
	UUG }	UCG }	UAG } Stop	UGG } Tryptophan (Trp)	G
C	CUU } Leucine (Leu)	CCU } Proline (Pro)	CAU } Histidine (His)	CGU } Arginine (Arg)	U
	CUC }	CCC }	CAC }	CGC }	C
	CUA }	CCA }	CAA }	CGA }	A
	CUG }	CCG }	CAG }	CGG }	G
A	AUU } Isoleucine (Ile)	ACU } Threonine (Thr)	AAU } Asparagine (Asn)	AGU } Serine (Ser)	U
	AUC }	ACC }	AAC }	AGC }	C
	AUA }	ACA }	AAA }	AGA }	A
	AUG } Methionine (Met)	ACG }	AAG }	AGG }	G
G	GUU } Valine (Val)	GCU } Alanine (Ala)	GAU } Aspartic acid (Asp)	GGU } Glycine (Gly)	U
	GUC }	GCC }	GAC }	GGC }	C
	GUA }	GCA }	GAA }	GGA }	A
	GUG }	GCG }	GAG }	GGG }	G

Fig: 2.1 Genetic code table (source references)

The number of permutations is 64 i.e. combinations. Any modification to these pairings could result in protein malfunction and, ultimately, a defective response condition. Any technique or technology that can read DNA sequences is required in order to diagnose, treat, or compare the DNA sequences.

Dna sequencing was firstly proposed by Sanger called first generation sequencing. Technologies which are used for DNA sequencing are

2.1 Sanger Sequencing: Sanger sequencing^[25], also known as chain termination sequencing, is a method of DNA sequencing that was developed by Frederick Sanger and his colleagues in the late 1970s. The DNA sequences of the provided complete genomes, including the human genome, have been determined using this popular technique for DNA sequencing. DNA polymerase is used in the Sanger sequencing technique to copy a single-stranded DNA template. Each nucleotide added to the expanding DNA strand by the DNA polymerase is marked with a fluorescent dye. Moreover, the DNA polymerase contains a modified nucleotide that, when added to the expanding DNA strand, stops DNA synthesis. The DNA fragments are sorted by size using gel electrophoresis after the DNA polymerase has integrated the tagged nucleotides. The fluorescent dyes make it possible to see the DNA fragments and determine the nucleotide order.)

2.2 Capillary Electrophoresis: (Proteins, nucleic acids, and carbohydrates are only a few examples of the many molecules that can be analyzed and quantified in biochemistry, chemistry, and molecular biology.

A sample is put into a tiny capillary tube that is filled with an electrolyte buffer solution to perform capillary electrophoresis^[19]. Charged molecules move through the buffer solution when an electric field is introduced across the capillary. The size, shape, and charge of the molecules separate them, with smaller, more strongly charged molecules travelling more swiftly than bigger, less strongly charged molecules. As the split molecules move past a detector at the capillary's end, they are picked up. A mass spectrometer, a fluorescence detector, or a UV-visible spectrophotometer can be used as the detector. The identity and number of the molecules in the sample are then determined by analyzing the data that has been generated.

2.3 Next Generation Sequencing: Sometimes referred to as high-throughput sequencing, this term describes a collection of cutting-edge DNA sequencing techniques that enable the quick and effective sequencing of substantial volumes of DNA or RNA. The discipline of genomics has been completely transformed by these technologies,^[26] which have been used to sequence the genomes of several creatures, including humans. Millions of DNA or RNA fragments are routinely sequenced at once using NGS technology' massively parallel sequencing method. This is achieved by combining techniques for sample preparation, building libraries, and sequencing techniques like Illumina, Ion Torrent, and Pacific Biosciences sequencing.

2.4 DNA Microarray Technology (DNA Chip):¹ Also known as gene expression profiling or DNA chip technology is an effective technique that enables researchers to investigate the expression of numerous genes in a single experiment. A tiny glass slide the size of a postage stamp that includes thousands of microscopic dots or probes that stand in for particular genes or DNA sequences makes up the technology. Before using a DNA microarray,^[17] researchers must first separate RNA from target cells or tissues and turn it into complementary DNA (cDNA). After being dyed with a fluorescent marker, the cDNA is subsequently hybridized to the probes on the microarray slide. Each location on the slide emits an amount of fluorescence that is proportional to the amount of cDNA that hybridizes to the probe, which in turn represents the level of the relevant gene's expression.

The collected information offers a snapshot of thousands of genes' levels of expression during a single trial. It can be used to determine whether genes are activated or deactivated in response to a specific medication, illness, or environmental factor.

Next-generation sequencing (NGS) is rapidly becoming popular due to its large-scale whole genome sequencing (WGS)^[4] accessibility, high sensitivity to discover low frequency alterations, and speedy and accurate comparisons.

III. Why NGS?

NGS has several benefits over conventional Sanger sequencing, including a higher throughput, quicker processing, and a cheaper cost per base. NGS has made it possible to analyze intricate genomic regions, including repetitive sequences and structural differences, and it has made it easier to find genetic changes and disease-causing mutations.^[2]

NGS has a wide range of uses in biotechnology, medicine, and research. To find mutations and to tailor cancer therapy, it is employed in cancer genomics. In addition, it is employed in the diagnosis of infectious diseases, the investigation of the micro biome, and other fields.

NGS has, in general, considerably improved our understanding of genomes and opened up a variety of new directions for investigation and innovation.

The burrows wheeler transform (BWT)-based mapping methods, such as burrows wheeler alignment (BWA) and bowtie tools, are the most frequently used algorithms in next-generation sequencing applications. When compared to traditional seed-based algorithms, these are said to be quick and precise. A whole genome sequencing (WGS) dataset can be mapped in approximately 3 hours using hardware acceleration using tools such as GPU (Graphical Processor Unit)^[12] or FPGA (Field-Programmable Gate Array), although this normally requires almost 24 hours i.e 1day of time without the use of hardware acceleration. Burrows wheel transform is quick but less reliable tool, whereas seed based technique is slow but more reliable.

IV. Proposed System

4.1 Using FANSe3 method

Thus, FANSe2 was generated to have both the benefits. It is based on real-world sequencing statistics to condense the mistake rates and considerably to speed up mapping process without sacrificing precision. FANSe3^[11] is a faster and more reliable method because FANSe2's error rate is not zero. It is both quick and precise. Without the requirement for any hardware acceleration features, it can read/map

4.1.1 a 30-human whole-genome sequencing (WGS)^[27] dataset in 30 minutes;

4.1.2 a 50-human whole exome sequencing (WES)^[10] dataset in 30 seconds; and

4.1.3 a typical mRNA-sequence dataset in seconds in a single-server node without the need for any hardware acceleration feature.

FANSe3 was created in the period from the year 2015 to 20. Its mismapping rate, as calculated analytically, is $1e-6$. To maximize the parallelizability, it is fully optimized for a cloud-based server infrastructure. The FANSe series mapping technique, which can map crores of sequencing reads to the referenced sequences, is in its third generation i.e FANSe3. This method is 40–200 times more quicker than FANSe2. Disk I/O is used to achieve this best performance primarily in applications which involves transcriptome and exome sequencing. Direct read trimming, effective indel support, and unmasked reference genome support are just a few of the strange features that FANSe3 has to deal with. FANSe3 is a bioinformatics software tool used for the analysis of next-generation sequencing (NGS) data. It is generally used to analyze RNA-seq (RNA sequencing) data, which is a method for figuring out how many genes are expressed at what levels in a biological sample. FANSe3 performs several important tasks in RNA-sequence analysis, including:

4.1.4 Quality control: Raw RNA-sequence data quality is checked by FANSe3, which can be impacted by things like sequencing mistakes and sample deterioration..

4.1.5 Read mapping: The RNA-sequence readings are mapped by FANSe3 to a reference transcriptome or genome. This makes it possible for scientists to pinpoint which genes are expressed in a particular sample.

4.1.6 Gene expression quantification: Gene expression levels in the RNA-sequence data are quantified by FANSe3. This gives crucial details about how various genes in the biological sample work.

4.2 File format

The FASTQ file format is used to hold this data, however GFF3 and BGEN have been more recently used.^[24] A DNA or RNA sequence and the related quality ratings are stored in a file format called Fastq. Each record contains four lines of text in this text-based format:

14.2.1 The first line starts with a '@' symbol and consists of a unique identifier for the sequence.

14.2.2 The second line contains the actual DNA or RNA sequence.

14.2.3 The third line starts with a '+' symbol and may consist of another identifier or be left empty.

14.2.4 The fourth line consists of the quality scores for the sequence in ASCII characters which is widely used in sequencing data and bioinformatics.

Large-scale genotype data, such as those which are collected from genome-wide association studies(GWAS)^[6], are stored in the binary BGEN file format. It is anticipated to be more effective than text-based formats, such VCF, in regard to read/write performance and storage space. Information about the samples, variations, and genotyping probabilities is present in each BGEN file. For our benefits to reduce file size, it uses a variety of compression algorithms. Many software programmes and libraries are used for manipulating and analyzing genotype data, which may also include PLINK, BGENIE, and Hail, that support the BGEN format.

V. Functionality

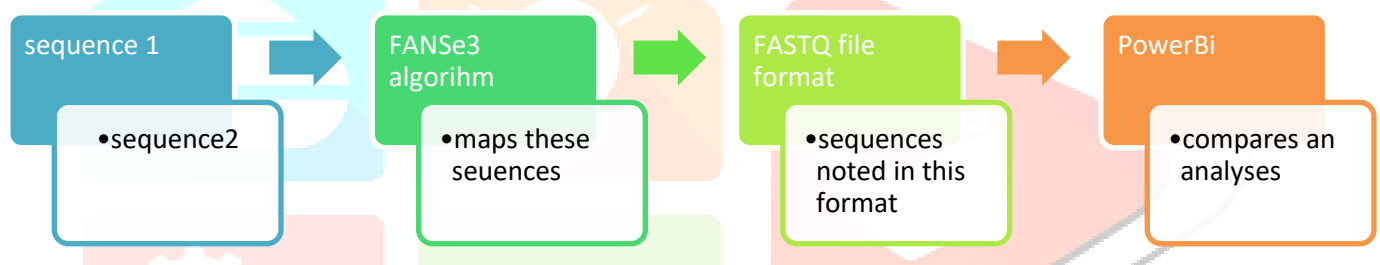
5.1 Using Power BI For Data Analysis:

Power BI^[14] can be a very useful tool for data analysis in sequencing^[13], as it allows users to connect to various data sources, such as sequence alignment files and variant call files, and perform data transformation and modeling. Power BI can be a useful tool for data analysis in sequencing. Charts, tables, and maps are just a small number of the built-in data visualization choices that Power BI offers, all of which can be used to explore and interpret sequencing data in a number of different ways. The capability of Power BI to manage enormous datasets and carry out data aggregation and filtering is one of the software's chief advantages when it comes to sequencing of data. In order to spot the patterns and trends in sequencing data^[1], such as differences in gene expression or the prevalence of particular variants, this can be helpful for analyzing. Additionally, Power BI can be used to create interactive dashboards that allow users to explore sequencing data in an intuitive and dynamic way.

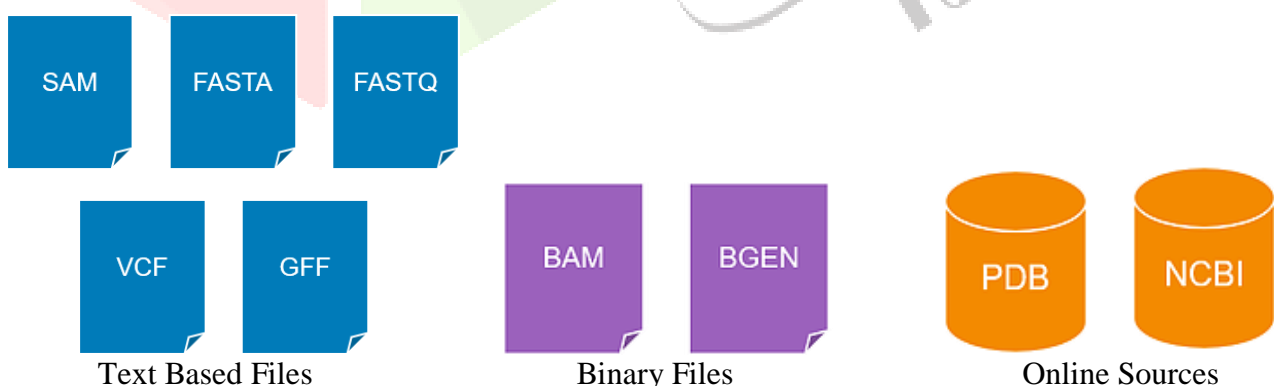
Some of the specific use cases in sequencing where Power BI can be useful include:

- Visualizing the coverage of sequencing data across a genome
- Analyzing the distribution of variants across samples
- Generating interactive plots of gene expression
- Creating reports and summary statistics of sequencing data

DNA sequences can be mapped by using FANSe3 algorithm of Next Generation Sequencing. The read sequences are to be compared with the reference sequences. This comparison can be done using Power BI.



Power BI is a collection of apps, software services, and connectors which can work together to turn your unconnected sources of data into visually immersive, coherent, and interactive insights. Power BI can work with files of FASTQ(FASTA) to BAM type. These data sources can be broken down into three main domains:



As there are no inbuilt connectors in genomics for Power BI, a blank query data editor is required. Generally power query reads the data in Power BI, which associates or connects to everything from CSV files to spark clusters.

5.1.1 Text Files: To analyse our files, M script is written in the query editor. For example SAM file is analysed as:

```
// Read SAM Files let // Read in file Source =
Table.FromColumns({ Lines.FromBinary(File.Contents("C:\Users\Colby\Documents\GitHub\bioPower
BI\bam_and_sam\sample.sam"), null, null, 65001)}), // Skip @ lines#"Filtered Rows" =
Table.SelectRows(Source, each not Text.StartsWith([Column1], "@")), // Split into columns by \t character
and assign names// Note: This removes and values past the 11 standard columns#"Split Column by
Delimiter" = Table.SplitColumn("#"Filtered Rows", "Column1", Splitter.SplitTextByDelimiter("#(tab)",
QuoteStyle.Csv),
{"QNAME", "FLAG", "RNAME", "POS", "MAPQ", "CIGAR", "RNEXT", "PNEXT", "TLEN", "SEQ", "QUAL"
}), // Change data types#"Changed Type" = Table.TransformColumnTypes("#"Split Column by Delimiter",)
in#"Changed Type".
```

5.1.2 Binary System:

These include file types of BGEN^[9] and BAM. These binary files^[20] can be read using python or R Language but Power BI makes it more simpler.

BGEN file is analysed as:

```
1 // Read BGEN Files
2 let
3 // Use rbgen in an R Script to get the data from the .bgen file
4 Source = R.Execute("library(rbgen)#(If)file <- "C:\\Users\\Colby Ford\\Desktop\\bioPower
5 BI\\bgen\\sample.bgen"#"(If)dataset <- as.data.frame(bgen.load(file))"),
6 sample = Source{[Name="dataset"]}[Value]
7 in sample This will be available as table format
```

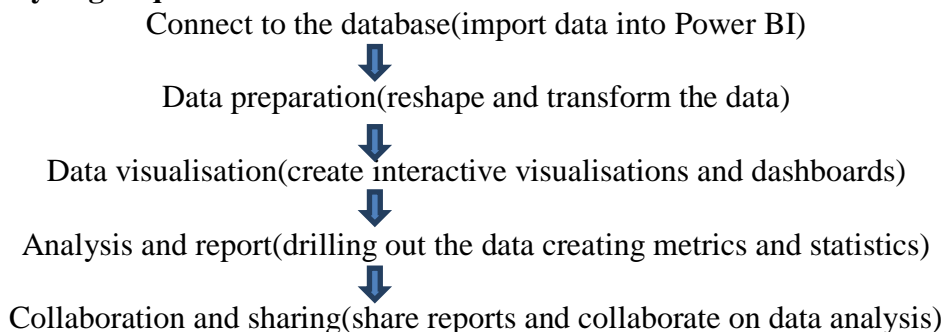
qname	flag	rname	mpos	mapq	cigar	isize	pos	mrm	qwidth	strand	seq	qual
SOLEXA-1GA-2_2_FC20EMB5:102:214:278	0	chr1	0	25	36M	0	10148		36	+	CCCTAACCCCTAACCCCTAACCCCTAACCCCTAAC	hbhhhxUYhT_ULZdLRtKNIM
SOLEXA-1GA-2_2_FC20EMB5:102:266:720	16	chr1	0	25	36M	0	875956		36	-	GGCCTTTAGACGCGGGCGTGCCTGGCATTGGGGC	A@FC@BBEGIOIEDBGGOLI
SOLEXA-1GA-2_2_FC20EMB5:103:913:330	16	chr1	0	25	36M	0	832906		36	-	AATATTTACAGCTAGCAGTCTCTTTTTTCCCTAG	LDDEDCThK@S_SVEOJWW
SOLEXA-1GA-2_2_FC20EMB5:105:57:684	0	chr1	0	25	36M	0	788329		36	+	GCACACACTCACTCACTATCTTTTTTTGGAGA	hhhhhhhhhhhhhhShgZheel
SOLEXA-1GA-2_2_FC20EMB5:105:653:895	0	chr1	0	25	36M	0	799485		36	+	GAGGGGGGGGAAATCCCTAACCCCTAACCC	hhhhhhhhhhhhhhShjT KUK
SOLEXA-1GA-2_2_FC20EMB5:108:485:455	16	chr1	0	25	36M	0	10156		36	-	CTAACCCCTAACCCCTAACCCCTAACCCCTAAC	CFVIUIIhGPROGhRhhhhhh
SOLEXA-1GA-2_2_FC20EMB5:108:987:867	16	chr1	0	25	36M	0	713372		36	-	TGCCCGGCAACATAGCAGGACCCCGCCGACCCA	?A>DC->?HE@GJAHGIGPHI
SOLEXA-1GA-2_2_FC20EMB5:109:305:261	0	chr1	0	25	36M	0	798366		36	+	GGTAACATGCTGTTCTGCTGGGTGAGGACCAT	hhhh^cRahXh_RhUhQJYZHI
SOLEXA-1GA-2_2_FC20EMB5:111:674:649	0	chr1	0	25	36M	0	757263		36	+	CCACCTGTCTACACTACCTGCCTCCCGAAGATC	hhhhhhchZbhShhhhhRhWZJe
SOLEXA-1GA-2_2_FC20EMB5:115:617:526	16	chr1	0	25	36M	0	872184		36	-	ACAGCTGAGAGGGCGTTGATAAAATTAATGCC	OGUKQJSLWTTaUMOVtgSh
SOLEXA-1GA-2_2_FC20EMB5:116:195:838	0	chr1	0	25	36M	0	762074		36	+	ACCCCACTACTGCTGCTCAGTCACTCTCTCAC	hhhhhhZyhhhhch XWLYSH
SOLEXA-1GA-2_2_FC20EMB5:116:56:969	0	chr1	0	25	36M	0	880361		36	+	CCCCCGCCACGTTGCTGCTCCCGCCAGATG]^TYRPQJMQGRKENMCDMB@
SOLEXA-1GA-2_2_FC20EMB5:116:646:499	16	chr1	0	25	36M	0	569895		36	-	TACTAACCAACACTAACCATATACCAATGATGC	>SFDhTUnhcYRMRhQshhhw
SOLEXA-1GA-2_2_FC20EMB5:116:932:930	0	chr1	0	25	36M	0	857815		36	+	AGACACAGACTCAGGAGGAGGACAGGAACTC	fhNhPZ IISYGNR HgNOIGLN
SOLEXA-1GA-2_2_FC20EMB5:117:1001:868	16	chr1	0	25	36M	0	864956		36	+	GCAGAAGGGTCAAAAGTACGGTGGGATGGCAGT	@>GBFGE>CKFKFHFFHJOK
SOLEXA-1GA-2_2_FC20EMB5:118:992:450	16	chr1	0	25	36M	0	848071		36	-	GCTGTTCTGGGGTGGGCGATGACCCCTCTGCC	EA?BHAEMCNMHHMHMYPF
SOLEXA-1GA-2_2_FC20EMB5:119:270:438	0	chr1	0	25	36M	0	880420		36	+	ACCCCAAGACCATCACCTCCGAGTTGCTCTTTC	fiQ^hhNagOVIIS^QCb@GTMI
SOLEXA-1GA-2_2_FC20EMB5:119:712:762	0	chr1	0	25	36M	0	883496		36	+	GGGAGGAGACACACCTGAGGACACAGCCGCAC	hhhhhhvdy_LLFJNLOHINIGD
SOLEXA-1GA-2_2_FC20EMB5:12:305:700	0	chr1	0	25	36M	0	807900		36	+	GGGAGGGGGTTTCATTGTTCCACTGCAGTGAGGT	hhhRyH^a^KYNNLNUJIMJEV

Fig: 5.1 BGEN file format sample (source references)

5.1.3 Online Sources:

Power BI makes it very easy to get the information or data from the web or internet. The reports produced from FANSe3 are analysed by Power BI to produce efficient data or it even helps to compare with the referenced sequences.

5.2 Power BI Analyzing Sequence:



VI. Reports:

6.1 Traditional records:

Traditional records [5] can refer to any type of records that are created and maintained in a physical format, such as paper, film, or other tangible media. Traditional records can include financial records, legal records, historical records, and many other types of documents that are used to document and preserve information. In the past, traditional records were the primary way of recording and preserving important information. However, as technology has advanced, many organizations have moved away from traditional records and towards digital records that can be more easily accessed, searched, and shared. Despite this shift towards digital records, traditional records still have an important role to play in many organizations. For example, many legal and financial records must be maintained in their original physical form in order to be admissible as evidence in court.

6.2 Traditional medical records:

Traditional medical records [16] typically refer to paper-based records that contain information about a patient's medical history, diagnosis, treatment, and other relevant medical information. These records are usually maintained by healthcare providers such as doctors, nurses, and hospitals. Traditional medical records are typically organized into sections that include the patient's personal information, medical history, clinical notes, laboratory results, radiology reports, and other pertinent medical data. These records are usually stored in file cabinets or other physical storage systems and require manual handling and updating. One of the challenges of traditional medical records is that they can be difficult to access and share among different healthcare providers, especially in cases where the patient has multiple medical providers or receives care in different locations. Additionally, traditional medical records can be vulnerable to loss, damage, or theft. Today, many healthcare providers are transitioning from traditional paper-based medical records to electronic medical records (EMRs)^[21] and electronic health records (EHRs)^[3], which offer a more efficient and secure way to manage and share patient information.

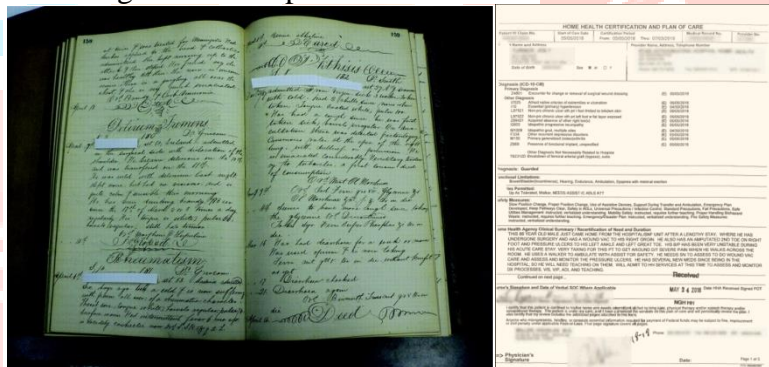


Fig: 6.2 Traditional Medical record Sample(source references)

6.3 Bi Report:

Some examples of traditional medical records include:

Medical history forms: These are documents that patients fill out to provide information about their medical history, including any previous illnesses, surgeries, allergies, and current medications.

Progress notes: These are written or typed notes that healthcare providers use to document a patient's diagnosis, treatment plan, and progress during each visit.

Radiology reports: These documents contain the results of medical imaging tests, such as X-rays, CT scans, and MRIs.

Laboratory reports: These documents contain the results of blood tests, urine tests, and other types of medical tests.

Prescription records: These documents contain information about a patient's current and past medications, including dosage and frequency.

Consultation reports: These are written reports that healthcare providers use to document consultations with other healthcare professionals, such as specialists or surgeons.

6.4 Advantages Of Using Power BI To Analyse Medical Reports:

Power BI is a powerful data visualization and business intelligence tool that can be used to analyze and gain insights from large and complex datasets, including medical records. Some advantages of using Power BI in analyzing medical reports include:

Faster insights: Power BI allows users to quickly and easily create interactive dashboards and reports that can provide real-time insights into medical data. This can help healthcare providers make faster and more informed decisions.

Customizable visualizations: Power BI provides a wide range of customizable data visualizations that can help healthcare providers better understand and analyze medical data. For example, charts, graphs, and maps can be used to show trends in patient data, identify high-risk patients, and monitor treatment outcomes.

Integration with other data sources: Power BI can be integrated with other data sources, such as electronic medical records (EMRs) and medical billing systems^[7]. This can help healthcare providers to access and analyze a wide range of medical data from different sources in a single dashboard.

Improved data accuracy: Power BI can help improve data accuracy by identifying data inconsistencies and errors in medical reports. This can help healthcare providers ensure that patient data is accurate and up-to-date.

Better communication and collaboration: Power BI allows healthcare providers to easily share medical reports and dashboards with other team members, such as physicians, nurses, and administrators. This can help improve communication and collaboration across different departments and improve patient care.

VII. Conclusion:

DNA sequencing is used to detect and treat diseases early, even before symptoms appear and the produced reports by FANSe3 are analysed by Power BI. It is done by using sanger sequencing, capillary electrophoresis, Next Generation Sequencing, DNA microarray technology,. The order of nucleotides is determined through DNA sequencing, which can vary in at least one way, causing a difference or disease. DNA sequencing has revolutionized drug development by enabling researchers to find new drug targets and evaluate a huge number of molecules. It is based on the four chemical bases, Adenine, guanine, thymine, and cytosine, which are joined with the help of a sugar and phosphate molecule. These bases create a unique genetic code, which has an impact on the body's ability to produce various chemicals and proteins. Any alteration to these sequences could cause the specific protein to breakdown. Technologies used for dna sequencing include

- Sanger sequencing,
- Capillary electrophoresis,
- DNA polymerase
- Next generation sequencing (NGS)

of all the techniques Next Generation Sequencing is mostly and widely used as it has several benefits over conventional Sanger sequencing, including a higher throughput, quicker processing, and a cheaper cost per base. The sequences are read by using FANSe3 and the reports are produced. The produced reports are analyzed by Power BI. Power BI also helps to interpret the data from traditional medical records so as to be user friendly.

VIII. Future Scope

FANSe3 and Power BI have a lot of potential for sequencing data analysis, as technology advances and more data is generated, both tools will become more powerful and user-friendly, thus making it more accessible to researchers, and providing more accurate and detailed functional annotation and visualization of sequencing data.

References:

- [1]An book on 'Understanding sequencing data as compositions: an outlook and review' by 'Thomas P Quinn'–<https://academic.oup.com/bioinformatics/article-abstract/34/16/2870/4956011>
- [2]A book on 'Targeted Next-Generation Sequencing can Replace Sanger Sequencing in Clinical Diagnostics' by 'Birgit Sikkema-Raddatz'– <https://onlinelibrary.wiley.com/doi/abs/10.1002/humu.22332>
- [3]An article on 'Electronic Health Records: Then, Now, and in the Future' by R. S. Evans–<https://www.thieme-connect.com/products/ejournals/html/10.15265/IYS-2016-s006>
- [4]An article on 'Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing' by 'Sang Tae Park and Jayoung Kim'–<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5169091/>
- [5]A book on 'Blockchain Technology Disrupting Traditional Records Systems' by 'Christoffer Koch'– https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2997588
- [6]A web reference on –<https://www.genome.gov/genetics-glossary/Genome-Wide-Association-Studies>
- [7]An article on 'Central Billing System for Personal Bills' by 'Yi Huang and Bin Wang'– <http://www.ijimt.org/papers/534-C016.pdf>
- [8]A web reference on – <https://jamanetwork.com/journals/jamaneurology/fullarticle/785995#:~:text=Background%20Myotonic%20dystrophy%20is%20caused,or%20even%20absent%20in%20others>
- [9]A web reference on –https://www.well.ox.ac.uk/~gav/bgen_format/
- [10]A book on 'Clinical application of whole-exome sequencing across clinical indications' by 'Kyle Retterer MS, Jane Juusola PhD Megan T. Cho ScM, Patrik Vitazka MD'– <https://www.nature.com/articles/gim2015148>
- [11]An article on 'The Ultrafast and Accurate Mapping Algorithm FANSe3: Mapping a Human Whole-Genome Sequencing Dataset Within 30 Minutes' by 'Gong Zhang'–<https://link.springer.com/article/10.1007/s43657-020-00008-5>
- [12]An article on 'Ecological Impact of Green Computing Using Graphical Processing Units in Molecular Dynamics Simulations' by 'Izabele Marquetti'–<https://www.igi-global.com/article/ecological-impact-of-green-computing-using-graphical-processing-units-in-molecular-dynamics-simulations/216906>
- [13]A web reference on –<https://sapac.illumina.com/informatics/sequencing-data-analysis/dna.html>
- [14]An article on 'Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data' by 'Louis T. Becker'– <https://www.tandfonline.com/doi/abs/10.1080/00987913.2019.1644891>
- [15]An article on 'Complete nucleotide sequences of two RNA segments of human picobirnavirus' by 'Mitsutaka Wakuda , Yaowapa Pongsuwanna'– <https://www.sciencedirect.com/science/article/abs/pii/S0166093405000637>
- [16]An article on 'Research on the rules of Traditional Chinese Medicine treatments of insomnia based on Ancient and Modern Medical Records Cloud Platform' by 'Wenhua ZHANG; Jing LI; Caifeng DU; Ziwei WANG; Caixia GUO; Yong ZHAO'. – <https://pesquisa.bvsalud.org/portal/resource/pt/wpr-882565>
- [17]A book on 'A REVIEW ON DNA MICROARRAY TECHNOLOGY' by 'Amandeep Singh, Naresh Kumar'–<https://www.academia.edu/download/33749084/45-1387896151.pdf>
- [18]A web reference on- https://www.google.com/url?sa=t&source=web&rct=j&url=https://my.clevelandclinic.org/health/diseases/21751-genetic-disorders&ved=2ahUKEwii25imzqT9AhWmtlYBHeLVDsQQFnoECA0QBQ&usq=AOvVaw217pa8_r3qbejXZXxSKYV
- [19]An article on 'Capillary Electrophoresis' by 'Rachel K. Harstad'–<https://pubs.acs.org/doi/pdf/10.1021/ac00179a002>
- [20]A book on 'Human Resource Analytics using Power Bi Visualization Tool' by 'Mohammed Ameer' – <https://ieeexplore.ieee.org/abstract/document/9120897/>
- [21]A book on 'Educational Impact of the Electronic Medical Record' By 'panel Paul J. Schenarts MD'– <https://www.sciencedirect.com/science/article/pii/S1931720411002984>
- [22]A book on 'The Genetic Code and RNA-Amino Acid Affinities' by 'Michael Yarus'–<https://www.mdpi.com/2075-1729/7/2/13>
- [23]An article on 'Structure and function of the human genome' by 'Peter F.R. Little'– <https://genome.cshlp.org/content/15/12/1759.short>
- [24]An article on 'an ultra-fast all-in-one FASTQ preprocessor' by 'Shifu Chen' – <https://academic.oup.com/bioinformatics/article-abstract/34/17/i884/5093234>

- [25]A web reference on-<https://www.cd-genomics.com/blog/sanger-sequencing-introduction-principle-and-protocol/>
- [26]An article on 'Overview of Next-Generation Sequencing Technologies' by 'Barton E. Slatko, Andrew F. Gardner' – <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpmb.59>
- [27]A book on'Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing' by 'Sang Tae Park and Jayoung Kim' – <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5169091/>
- [28]A web reference on –shorturl.at/dqrFU
- [29]A web reference on –https://www.google.com/url?sa=t&source=web&rct=j&url=https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-data-sources&ved=2ahUKEwix27Of9qX9AhUHumMGHStCDREQFnoECA0QAQ&usg=AOvVaw32ulTw33NhWHyuVK8xoUH_
- [30]A web reference on-<https://www.ncbi.nlm.nih.gov/books/NBK26821/>

BIBLIOGRAPHY



Miss. Kurumaddala Veda Srija Mahalaxmi: is studying her pre final year Bachelor of Pharmacy in Srinivasa Rao College of Pharmacy. With her interest in Bioinformatics, Databases she attended and won in various international conferences on the above topic. This article have been evolved from an idea to understand the flaws in conventional reporting and keeping time consistency, quality report generation, comparative analysis in FANSe3 using Power BI.



Kandhati Tulasi Krishna Kumar Nainar: Training & Placement Officer with decade plus experience in training & placing the students into IT, ITES & Core profiles & trained more than 9,000 UG, PG candidates & trained more than 350 faculty through FDPs. Authored various books for the benefit of the diploma, pharmacy, engineering & pure science graduating students. He is a Certified Campus Recruitment Trainer from JNTUA, did his Master of Technology degree in CSE from VTA and in process of his Doctoral research. He is a professional in Pro-E, CNC certified by CITD He is recognized as an editorial member of IJIT (International Journal for Information Technology & member in IAAC, IEEE, MISTE, IAENG, ISOC, ISQEM, and SDIWC. He published articles in various international journals on Databases, Software Engineering, Human Resource Management and Campus Recruitment & Training.