An Overview of Next Generation Sequencing and its Application in Neurodegenerative Diseases

Kritika Sehgal, Stuti Agarwal, Aarushi Tandon, Avantika Mishra,*Chakresh Kumar Jain,
Department of Biotechnology
Jaypee Institute of Information Technology,
Noida, UP
India, 201309

Abstract— Sanger sequencing has helped us lay the foundation for Next Generation Sequencing (NGS). Neurodegenerative diseases are least curative in nature when compared to others and the only measures that can be taken are at the onset stages which are easy to treat. This is where Next Generation Sequencing comes into play.

This paper discusses four randomly selected neurodegenerative diseases namely; Alzheimer’s, Parkinson’s, Ataxia and Multiple Sclerosis, their respective mutated genes, their possible causes and treatment. NGS technologies have been used to study and detect these diseases, which are discussed later. With the advent of NGS, it has been possible to sequence and study genomes at a much faster rate and in a much shorter span of time. The scope of NGS and the basis of its functionality have been thoroughly deliberated.

Keywords: NGS, Neurodegenerative Diseases, Deep sequencing, Biomarkers

I. INTRODUCTION

There has been a sea change since the beginning of the two dimensional approach towards DNA Sequencing. There were several reasons for switching to a high throughput sequencing method called the Next-Generation Sequencing (NGS), two of which were high cost and time consumption. Sequencing is done to find mutated alleles causing pathological diseases by comparing the DNA sequence of a healthy individual and that of a diseased person which aided in customized drug design, since each individual has a unique genome. NGS therefore helps characterize diseases on epigenetic, genomic and transcriptomic levels. Parallel sequencing can be used as means of transcriptomic analysis of mRNAs, small RNAs, non-coding mRNAs, etc. [1]. Sanger sequencing, the predecessor to Next-Generation Sequencing laid the ground for Next-Generation Sequencing.

Sanger Sequencing employed an enzyme-based approach which was advantageous over the Maxam Gilbert Sequencing method since it used chemicals and radioisotopes which were toxic in nature [2]. Sanger Sequencing, also called the “Chain Termination method”, works on the principle of selective integration of chain terminating ddNTPs (dideoxynucleotide phosphate) with the help of DNA Polymerase enzyme during DNA replication. The absence of the 3'OH group of ddNTPs results in chain termination, which would otherwise form a phosphodiester bonds (required for the chain elongation). For detection in automated machines, these ddNTPs are fluorescently labeled. However, only 800 base pairs can be sequenced using Sanger Sequencing [3].

Next Generation Sequencing can be used to determine the genotypes that explain phenotypes, consequential in voluminous amount of DNA information. This demand cannot be fulfilled by the 2nd generation method i.e., Sanger Sequencing, but NGS, which is a much faster and accurate process of sequencing, can be engaged [4].

Next Generation Sequencing technology functions on platforms including Ion Torrent and Illumina methods. Ion Torrent method harnesses the power of semiconductor technology and detects the protons released when a nucleotide is incorporated during synthesis. Illumina method adopts a sequencing-by-synthesis approach, utilizing fluorescently labeled reversible-terminator nucleotides, on clonally amplified DNA templates immobilized to an acrylamide coating on the surface of a glass flow cell [5].

II. SCOPE OF NEXT GENERATION SEQUENCING

Next Generation Sequencing is a promising and upcoming field in bioinformatics with the rising demand of cost-effective and high throughput sequencing. With the advent of New Generation Sequencing techniques, ranging from Illumina/Solexa, SOLID/Roche to Helicos, unparalleled chances for high-throughput genome analysis have surfaced. These techniques have been applied in varied contexts, some of which are non-coding RNA expression profiling, whole-genome sequencing, finding transcription factor binding sites, targeted re-sequencing, and the list goes on [6]. These technologies have shown a huge impact on metagenomics as well [7].

The application of NGS is not restricted to genomics, but it also finds application in transcriptomics and epigenomics. RNA-Seq which is a transcriptome profiling tool makes use of deep sequencing technologies [8]. ChIP-Seq, MBD-Se, MeDIP-Seq, and MeDIP-Seq are few tools used deep sequencing technologies [9][10]. Table 1 explains the contrast amongst different NGS platforms. The 454 Roche technology [11] is the first NGS technology which is commercially available and has successfully overcome the cloning requirement by using emulsion PCR [12]. Pyrosequencing approach [13][14] follows a pattern of sequencing after synthesis, by measuring inorganic pyrophosphate which is released, by chemiluminescence. Currently, the 454 platform can generate

IJERTV6IS020059 www.ijert.org
(This work is licensed under a Creative Commons Attribution 4.0 International License.)
Neurodegeneration can be originated at many different inflammation, leading to cellular apoptosis factors, oxidative stress, etc. Noxiousness of free radicals imbalanced defense mechanism of antioxidants, systemic levels. Several neurodegenerative diseases were stages of neuronal circuitry oscillating from molecular to classified as proteopathies as they were associated with diseases [20]. There can be a variety of conditions that lead to atrophy in neural cells is responsible for neurodegenerative of nerve cells cause dementia or ataxia. Suppression of basal death disease-diagnostic may be employed with the development of level and hence are correlated. They are incurable as of now, do they reproduce; instead they age over time and ultimately cells in the body, they don’t get replaced in the brain, neither do they reproduce; instead they age over time and ultimately die [24].

All these diseases are related to each other on the sub-cellular level and hence are correlated. They are incurable as of now, but with time, NGS technologies and their application in disease-diagnostic may be employed with the development of computational, experimental and technical platforms. Death of nerve cells cause dementia or ataxia. Suppression of basal atrophy in neural cells is responsible for neurodegenerative diseases [20]. There can be a variety of conditions that lead to dementia or ataxia - including death of nerve cells, imbalanced defense mechanism of antioxidants, overproduction of free radicals from environment, genetic factors, oxidative stress, etc. Noxiousness of free radicals underwrites to protein and DNA injury, tissue damage, inflammation, leading to cellular apoptosis [21]. Neurodegeneration can be originated at many different stages of neuronal circuitry oscillating from molecular to systemic levels. Several neurodegenerative diseases were classified as proteopathies as they were associated with aggregation of misfolded proteins [22]. Symptoms included disruption in emotional, cognitive and social behaviour, apathy, anxiety, euphoric mood, dysphoric mood and disinhibition [23]. Table 2 states the etiology of various neurodegenerative diseases.

Brain diseases have emerged as the leading contributors to global disease liability; for example, according to the (2004) WHO survey, the US National Institute of Mental Health (NIMH) approximates that about 1 in 4 American adults undergo mental disorders, with nearly 6% suffering serious infirmities as a result [24]. Cognitive disturbance (dementia) involves the degeneration of cerebral cortex, e.g., Alzheimer’s disease and Pick Disease. Movement diseases include degeneration of motor neurons, cerebellum and connecting tracts such as substantia nigra, basal ganglia (Parkinson’s and progressive supranuclear palsy) and multiple system atrophy [25]. With the increasing toll of neurodegenerative diseases on life, it is important to work out on techniques and methodologies to cure the same. The first and foremost step is the detection of causative agents which is done by using Next Generation Sequencing (NGS). As research advances, many similarities show that these ailments relate to one another on a sub-cellular level. Ascertaining these similarities offers scope for therapeutic advances that could improve many diseases diagnostics instantaneously [26]. Table 2 explains origin and causes of neurodegenerative diseases.

### Table 1 - Comparison between various NGS platforms [15]

<table>
<thead>
<tr>
<th>Technology</th>
<th>Principle</th>
<th>Analysis time</th>
<th>Average length read</th>
<th>Throughput (Mb/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roche titanium</td>
<td>Pyrosequencing</td>
<td>4 hours</td>
<td>300-500bp</td>
<td>20-30</td>
</tr>
<tr>
<td>Illumina MiSeq</td>
<td>Bridge amplification</td>
<td>27 hours</td>
<td>Upto 150bp</td>
<td>20</td>
</tr>
<tr>
<td>Ion Torrent</td>
<td>Incorporation of H’ ion</td>
<td>2 hours</td>
<td>200bp</td>
<td>320</td>
</tr>
</tbody>
</table>

### Table 2- Etiology of neurodegenerative disease [27]

<table>
<thead>
<tr>
<th>Condition</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein misfolding and/or defective degradation</td>
<td>Disruption of cellular/axonal transport</td>
</tr>
<tr>
<td>Oxidative stress and formation of free radicals</td>
<td>Actions and mutations of molecular chaperones</td>
</tr>
<tr>
<td>Mitochondrial dysfunctions</td>
<td>Dysfunction of neurotrophins</td>
</tr>
<tr>
<td>Fragmentation of neuronal Golgi apparatus</td>
<td>Neuro-immune processes</td>
</tr>
</tbody>
</table>

### III. NEURODEGENERATIVE DISEASES

Neurodegenerative Diseases distress both the central nervous system and the Peripheral Nervous System. It is a broad term, comprising of a number of conditions related to gradual loss of neurons and synapses in the Nervous System. Neurons are the fundamental unit of the nervous system. Unlike other cells in the body, they don’t get replaced in the brain, neither do they reproduce; instead they age over time and ultimately die [24].

All these diseases are related to each other on the sub-cellular level and hence are correlated. They are incurable as of now, but with time, NGS technologies and their application in disease-diagnostic may be employed with the development of computational, experimental and technical platforms. Death of nerve cells cause dementia or ataxia. Suppression of basal atrophy in neural cells is responsible for neurodegenerative diseases [20]. There can be a variety of conditions that lead to dementia or ataxia - including death of nerve cells, imbalanced defense mechanism of antioxidants, overproduction of free radicals from environment, genetic factors, oxidative stress, etc. Noxiousness of free radicals underwrites to protein and DNA injury, tissue damage, inflammation, leading to cellular apoptosis [21]. Neurodegeneration can be originated at many different stages of neuronal circuitry oscillating from molecular to systemic levels. Several neurodegenerative diseases were classified as proteopathies as they were associated with aggregation of misfolded proteins [22]. Symptoms included disruption in emotional, cognitive and social behaviour, apathy, anxiety, euphoric mood, dysphoric mood and disinhibition [23]. Table 2 states the etiology of various neurodegenerative diseases.

Brain diseases have emerged as the leading contributors to global disease liability; for example, according to the (2004) WHO survey, the US National Institute of Mental Health (NIMH) approximates that about 1 in 4 American adults undergo mental disorders, with nearly 6% suffering serious infirmities as a result [24]. Cognitive disturbance (dementia) involves the degeneration of cerebral cortex, e.g., Alzheimer’s disease and Pick Disease. Movement diseases include degeneration of motor neurons, cerebellum and connecting tracts such as substantia nigra, basal ganglia (Parkinson’s and progressive supranuclear palsy) and multiple system atrophy [25]. With the increasing toll of neurodegenerative diseases on life, it is important to work out on techniques and methodologies to cure the same. The first and foremost step is the detection of causative agents which is done by using Next Generation Sequencing (NGS). As research advances, many similarities show that these ailments relate to one another on a sub-cellular level. Ascertaining these similarities offers scope for therapeutic advances that could improve many diseases diagnostics instantaneously [26]. Table 2 explains origin and causes of neurodegenerative diseases.

1. **Alzheimer’s disease**

Alzheimer’s is a widespread functional disease of the human nervous system. There are a lot of theories suggesting the cause of Alzheimer’s disease (AD). It is predicted that AD might affect as many as 1 in 85 people globally by 2050 [28]. A few of them suggest that aluminum accumulation or abnormal protein formation may be the causative agents, while others draw their conclusion from studies suggesting imbalances in brain function and others are based on reports similar to that of other infectious diseases [29]. AD is caused by a variety of factors including environmental and genetic factors, age being the most common [30]. Risk factors include smoking, hypertension, depression, heart disease, arthritis and diabetes [31][32]. Additionally, exercise [33] and Mediterranean diet [34][35] can also prevent the onset of AD.

The pathophysiological process of AD begins well in advance to the diagnosis of dementia. Majority of AD patients do not show any symptoms in pre-clinical stages of pathological processes (estimated to be approximately 17 years) [36]. Therefore, early diagnosis is essential for therapeutic treatment or administration of disease modifying drugs [37].
1.1 NGS using biomarkers

Owing to Genome-Wide Association Studies (GWAS) done for various neurological diseases, many genes related to these diseases have been discovered. It provides information about the various genetic causes that lead to common diseases. Recent development in two main fields of GWAS and NGS technology, have helped in understanding the genetic causes of these diseases [36]. For rare Mendelian diseases, NGS can detect presence of novel genes that have mutations responsible for the phenotype. Scattered diseases are susceptible to GWAS, whereas those which can be traced on a family tree signifying Mendelian disease are better analyzed by NGS-based studies [38]. Hence, with the development of NGS, whole-exome sequencing (WES), and whole-genome sequencing (WGS), have become quicker and cheaper over the past few years [39]. It has been realized that genetic factors play a crucial role in development of AD [40]. Mutations in genes like APP, PSEN1 and PSEN2 are inherited in accordance with Mendel’s laws and directly cause early-onset AD (EOAD). However, many EOAD cases have shown that genetic factors except those mentioned above are responsible for pathogenesis of EOAD. In past few years, NGS is being widely used to discover such factors in small families which have unsolved EOAD. Advancement in NGS has helped in overcoming the drawbacks of GWAS and has substantially supported the hypothesis suggesting that rare variation can explain few of the genetic heritability in AD [41].

AD can be detected clinically by the presence of tangles and plaques which are insoluble and composed of beta-amyloid (Aβ), formed by sequential amyloid precursor protein proteolysis and hyperphosphorylated Tau proteins [42]. Use of markers like Aβ and tau diagnostic tools is under investigation for a long time now [43][44]. Persistent efforts have been made to develop molecular markers for diagnostic purposes that are easily accessible, have high specificity and are cheap [45].

1.2 Deep sequencing based on miRNA profiling

Many biomarker targets are currently employed for AD diagnosis. Micro-RNAs (miRNA) are a class of non-coding RNAs of approximately 22 nucleotides in length, and regulate post-translational transcription. Expression profiling of miRNA level have come across as a new class of substantial biomarkers that are currently being examined for the diagnosis of various diseases [46]. The mature miRNA is incorporated into RNA-induced-silencing complex (RISC), which binds to complementary sites in 3'UTR (untranslated region) of mRNA targets hence down-regulating gene expression [47]. They can be secreted into biological fluids and profiling can be done using few methods like quantitative real-time PCR (qRT-PCR), microarrays, and recently developed deep sequencing technologies [48]. The highest expression of tissue specific miRNA in found in the brain [49]. A number of deregulated miRNA have been discovered to be related to AD, some of them being miR-9, miR-20a, and miR-13 [50].

Despite the high RNase activity, microRNAs are generally protected from degradation due to their binding with RNA binding proteins, e.g. Lipoproteins [51]. In AD, microRNAs can possibly transport through blood-brain barrier (BBB) through thinning and perforations in vascular membrane [52]. Also, micro vesicles and exosomes can play an important role as carriers of microRNA across BBB, hence facilitating interaction between brain and distant organs through biological fluids [53].

The cellular components of blood provide abundant source of RNA species used for biomarker analysis. Most abundantly found in white blood cells (WBCs), miRNA analysis may provide an insight to the indirect causes of neurodegeneration or pathogenesis of sporadic AD [54].

Generally analysis of plasma and serum is used to detect miRNA profiles specific to a disease. However, in AD, a very few patients have miRNA biomarkers profiles. In tested patients, a down regulation of brain-enriched miRNAs was seen [55]. MiRNA detection in blood has been seen as advantageous for early diagnosis; however differential miRNA expression may not accurately represent deregulation in neuronal tissues related to neurodegenerative disease. CSF is a better source for diagnostic purposes of CNS (central nervous system) [56]. Studies involving CSF have mostly been a combination of miRNA microarrays [57], multiplex miRNA qPCR approach or target candidate miRNA approach [58].

<table>
<thead>
<tr>
<th>Sequencing</th>
<th>Alignment</th>
<th>Expression Profiling</th>
<th>Final Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample isolation from AD patient</td>
<td>Filtering the sequence</td>
<td>Quantification of expression value</td>
<td>Expression testing</td>
</tr>
<tr>
<td>Isolation of miRNA</td>
<td>Sequence mapping</td>
<td>Normalisation</td>
<td>Cross-check with sample</td>
</tr>
<tr>
<td>Preparation of sample for profiling</td>
<td>miRNA expression</td>
<td>Study of identified miRNA</td>
<td></td>
</tr>
<tr>
<td>Library preparation</td>
<td>Identification of dysfunctional miRNA</td>
<td>miRNA analysis related to disease</td>
<td></td>
</tr>
<tr>
<td>Screening of target miRNA</td>
<td>Identification of biomarker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequencing of selected miRNA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In order to implement miRNA deep sequencing in clinical practices, a standard order needs to be defined as in Table 3 [59] normally workflow involves:

Table 3- Steps in miRNA deep sequencing [59]

```markdown
<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Sample isolation from AD patient followed by miRNA profiling and library preparation.</td>
</tr>
<tr>
<td>2.</td>
<td>Library preparation, identification of dysfunctional miRNA and miRNA analysis related to disease.</td>
</tr>
<tr>
<td>4.</td>
<td>Sequencing of selected miRNA.</td>
</tr>
</tbody>
</table>
```

Various tools and scripts such as Perl are available for processing huge amounts of data by a streamlined workflow. [57] The technique involves collection of sample from the patient followed by miRNA profiling and library preparation. Identified disease causing agents i.e., miRNA are then sequenced [59].

2. Parkinson’s Disease

Parkinson’s disease is among the most common diseases prevailing amongst the elder and middle-aged population. It is a movement disorder marked by degeneration of neurons in areas of basal ganglia and deficiency of dopamine (neurotransmitter), the cause of which remains unknown. It can be categorized into primary or secondary level.
Primary Parkinson may be genetic or idiopathic while secondary Parkinson is mainly caused due to toxins. Accumulation of Lewy bodies in neurons of the mid-brain hampers dopamine-production causing the disease. Causes can be demarcated clinically and etiologically, including idiopathic paralysis agitans, encephalitis, cerebral atrophy or tumor, severe cerebral trauma, neurosyphilis, carbon monoxide and manganese poisoning [60]. Major cell death takes place in substantia nigra of the brain, especially the ventral part of pars compacta. Disruption of motor, ocular-motor, associative, limbic and orbitofrontal take place lead to the disease. Mostly, Parkinson’s disease is classified as a movement disease but it may show symptoms of sensory defects, sleep problems or cognitive difficulties as well. Major difficulty lies in recalling of learned information. Mutation in genes causes α-synuclein (SNCA) proteins to misfold [61]. Aggregation of SNCA causes communication loss between neurons leading to their gradual death [62][63]. MicroRNAs have been identified as regulators for development of pathological state of Parkinson’s disease with their validation through NGS in process [64]. Mutations in some genes have found to be conclusive which cause Parkinson’s disease. These genes code for α-synuclein (SNCA), parkin (PRKN), leucine-rich repeat kinase2, PTEN-induced putative kinase1 or PINK, ATP13A2 and DJ1[65]. In case of Sporadic Parkinson’s disease, SNCA and LRRK1 have been identified as risk factors. GWAS i.e. Genome Wide Association Studies are being explored extensively to categorize mutant alleles which give test positive for the disease [66]. Parkinson’s disease has increased the universal mortality rate by a great number and is a chief donor to the Global Burden of Diseases. The number of new cases is 60,000 every year with 10,3000 deaths were reported in 2013. Currently 1% of people above 65 years of age suffer from this disease [67]. Treatments for motor symptoms include medications like antiparkinson Levodopa-Carbidopa, MAO-B inhibitors and dopamine agonists (bromocriptine, pergolide, pramipexole). Cystathalamectomy help improve early symptoms [68]. Non-motor symptoms are treated by targeting autonomic, sleep and cognitive disturbances [69]. Surgery including deep brain surgery and ablative procedures become inevitable at advanced stage. The disease can be diagnosed by a neurological exam or by medical history. However, no lab tests are available which clearly identifies the disease. During an experiment conducted with the help of Next Generation Sequencing (NGS), genes were studied and novel and annotated variants were discovered in 237 unrelated Chinese patients. The over or under expression of these genes was responsible for Parkinson’s disease. Target gene capture technology was used for the detection of rare variants. 48 genes were enriched with Hi-Sequence-2000 producing high quality reads. Target Capture Region was 56.35% out of which 94% of yielded clean reads were mapped to reference genome. Four novel and six annotated non-synonymous Single Nucleotide Polymorphisms (SNPs) were found and validated using Sanger Sequencing. The aim of the experiment was to discover new mutations in genes using NGS by scanning of the entire genome which caused the disease [70].

3. Ataxia Disease
Ataxia Disease is the abnormality of the parts of nervous system (such as cerebellum) that are responsible for the coordination of other body parts. Causes for ataxia may be focal lesions, exogenous substances (such as antiepileptic drugs, ketamine and dextromethorphan), radiation poisoning, vitamin B12 deficiency and hyperthyroidism. Ataxia can be broadly classified as Cerebellar ataxia, Sensory ataxia and Vestibular ataxia [71][72]. When studied about the Alsace region in France, where mean age for the onset of a disease was 17 years, reported a pervasiveness rate of autosomal recessive cerebellar ataxia was found to be 5.3 in 1 lakh during 2002 to 2008 [73]. When studied from a Portugal report a pervasiveness of hereditary cerebellar ataxia was 4.4 in 100,000 made up of a conjecture recessive pattern of inheritance. Autosomal dominant pattern of inheritance was 0.8, congenital was 0.4 in 1 lakh and mitochondrial was 0.8[74]. Friedrich ataxia although remained the most trivial diagnosis of advancing hereditary ataxia, with a pervasiveness of 2-4 in 1 lakh in Caucasians [75][76]. When studied in Sweden, the coarse pervasiveness rate of ataxic cerebral palsy was 10.9 in 1 lakh live births[77]. The pervasiveness of non-progressive ataxia as a result from prenatal events, after eliminating patients with spasticity, was 13 in 1,00,000 in 6-22 year-old patients during 1971 to 1986 [78].

The development of Next-Generation Clinical Exome Sequencing (CES) has made it possible to perform genetic diagnosis for ataxia patients as part of a detailed clinical workup[79][80]. This is a potential tool adding effectiveness to the physician’s armory because, number of rare genes related to ataxia, over and extensive use of low-yield single-gene and genetic panel testing serves a convincing cost to patient’s health care[81]. The predictable widespread benefits of CES have already provoked recommendations for its incorporation as part of routine clinical algorithms[82][83]. Although CES is less expensive than performing traditional Sanger Sequencing for multiple single genes, there is finite data to support the comprehensive use of this testing as part to examine the patients with cerebellar ataxia. Of the 16 cases taken in a study with pathogenic variants of ataxia, most of them presented sporadically (69%, 11 of 16) and early-onset cases (63%, 10 of 16)[84]. 14 cases had autosomal recessive inheritance while 2 displayed autosomal dominant inheritance. Two recessive genes were found in more than one person namely, SYNE1 and SPG7. It was also observed that 16 another variants that were novel across 13 disease genes, gave a clear illustration of the advantages of CES in patients with heterogeneous phenotypes [85].

4. Multiple Sclerosis Disease
Multiple Sclerosis (MS) is a chronic inflammatory disease of the Central Nervous System (CNS) which has demyelinating lesions which are inflammatory [98]. It is assumed that MS affects the white matter primarily but it affects the grey matter as well. Genome wide association studies (GWAS) is used for studying the human genome.
In recent reports, GWAS was shown to have associations with single locus with the help of pathway and network based analysis. GWAS works on a simple ‘one SNP at a time’ approach, thus reducing the complexity of the complex diseases that occur commonly [99]. The genetic susceptibility variants of these complex diseases can be detected by GWAS. An association of MS has also been seen with the human leukocyte antigen (HLA) of the CNS. TKY2 gene is a functional variant which has proved to be protective of MS which is located on the chromosome 19p13 and encodes for a proline to alanine substitution in exon 21[100] [101] [102]. Magnetic Resonance Imaging is the conventional method used for diagnosing MS [103][104]. An experiment was conducted for -in vivo detection of cortical lesions using MRI at 3 Tesla. Basically, the impact of high magnetic field strength was to be investigated on the detection of rate of cortical lesions in MS by using a double inversion recovery (DIR) pulse sequence. The conventional MRI is always done at 1.5 Tesla but here DIR was now being used at Tesla. Another experiment was conducted, where the highest density of CD3- positive T cells were found in MS white matter lesions and the lowest number were detected in intracortical demyelinated lesions. This was equal to the lymphocyte density in non-demyelinated cerebral cortex within the same tissue block or cerebral cortex in control brains [105]. There were also some focal lesions with higher signal intensity such as intracortical lesions and mixed white and grey matters [106]. These intracortical lesions did not involve white matter but were purely in the cortex, whereas, the white matter-grey matter lesions were the ones located in the white matter as well as in the cortical grey matter. However, there was a problem with the conventional MRI technique earlier in the detection of cortical lesions because of the low sensitivity of it in grey matter manifestations. DIR sequences not only suppress CSF (cerebrospinal fluid) but also suppress the white matter, thus increasing the detectability of cortical lesions in MS significantly. The conclusion of the above experiment was that DIR brain MRI at 3 Tesla improves the sensitivity of the detection of cortical lesions compared to 1.5 Tesla [107][108].

These experiments became possible because of the use of Next Generation Sequencing which helped the scientists and researchers to find out the sequences in a much shorter time as compared to earlier. MS is said to be an autoimmune disease of the CNS mediated by T-cell which is responsible for a myelin antigen [109]. It is also found that MS is a two staged disease; one is the inflammatory phase and the other is the neurodegenerative phase. For better understanding of these phases specific therapeutic targets were developed [110]. MS usually begins at the onset of early maturity with an autoimmune inflammation response against components of the myelin sheath. And hence, it needs to be diagnosed as early as possible and next generation sequence is a step closer to such a diagnosis [111].

Multiple Sclerosis often has certain side effects apart from the symptoms of the disease itself such as fatigue and depression which affected the patients directly and indirectly mentally as well as physically[112][113]. The assessment and evaluation of neurodegenerative diseases have been detailed in table 4.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Genes</th>
<th>Symptoms</th>
<th>People affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s</td>
<td>FAD, apolipoprotein E (APOE), two proenin (PSEN-1 and PSEN-2)</td>
<td>Difficulty to retain information, apathy and exhaustion on, sun downing.</td>
<td>46.8 million by 2015</td>
</tr>
<tr>
<td>Parkinson’s</td>
<td>PINK1 (PARK6)</td>
<td>Resting tremor, Bradykinesia, Postural instability, Micrographic.</td>
<td>7-10 million by 2015</td>
</tr>
<tr>
<td>Ataxia</td>
<td>FXN, ATM, Abch7</td>
<td>Impaired limb coordination, Cognitive and Mood Problems, Fine motor incoordination. Tingling, numbness, blurred vision, muscle stiffness.</td>
<td>-</td>
</tr>
<tr>
<td>Multiple Sclerosis</td>
<td>IL7R, IL2R, CLEC1 6A, CD226 , TKY2</td>
<td></td>
<td>2.5 million by 2015</td>
</tr>
</tbody>
</table>

Table 4- Assessment of various diseases [114][115][116].

IV. COMPUTATIONAL TOOLS

A variety of tools and pipelines are available for different applications and analysis, for eg: Cufflinks, Bowtie, FastQC, StringTie, etc. However, there is no optimal tool for which RNA-seq can be applied. For instance, sequence genome quantification has to be achieved by assembling the reads into contigs followed by mapping them on the transcriptome for organisms not having genome information in databases. On the contrary, for organisms having genome sequence available in databases, transcripts can be identified by simply mapping reads from RNA-seq on the genome. [117] Different tools are required at different steps in RNA-seq including obtaining raw data, quantification and read alignment.

FastQC[118] is tool performed on Illumina Platform, whereas NGSQC[119] is another tool which can be performed to any platform. Both these tools perform the same function of analysis of GC content, Sequence quality, sequencing errors in duplicated reads, presence of adaptors, PCR contamination etc.

dbOrtho is a useful tool for orthologous comparisions. This tool is used for ortholog conversion of identifier from one species to identifier in different species. Identifier sequence types can be varied in input and output sequences.[120] FASTX-Toolkit and Trimmomatic are software tools involved in discarding low-quality reads and poor quality bases and also in trimming adaptor sequences. In general, mappability of the bases should be improved by removing bases which are too low in reads as the read quality deteriorated towards the 3’ end of the reads. Cufflinks[121][122] is another tool which estimates gene length in samples where gene length between samples cannot be ignored. It is also known for estimating transcript expression from mapping to the genome. It also uses GTF information which is very important in identifying expressed transcripts or it can also infer transcripts de novo from the mapping data alone as well.
TopHat use an expectation-maximization approach to estimate transcript abundances. However, this approach is accountable for biases like non-uniform read distribution along gene length. MiRscan is based on RNAfold formed from hairpin structures and can search for conserved sequences present in intragenic regions. After sequence identification, it compares the output to known miRNA attributes [123]. Multi-mapping reads among transcript and also the output can be allocated within the sample normalized values for sequencing biases. This is achieved by certain algorithm tools such as RSEM(RNA-Seq by Expectation Maximization)[124], eXpress[125], Sailfish[126], and kallisto[127]. We also need tools for estimating transcription estimation and NURD[128] does this from SE reads with a low memory and computing cost. Certain predictions can be refined by estimating or testing for associations between genes, miRNAs, pathways etc or by looking upon the relatedness or non-relatedness of both the targeted genes and the associated miRNAs. Tools such as CORNA [129], MIMA[130,131] and SePIA[132] helps in these predictions. miRTRAP is a systematic tool used for the identification purposes of specific miRNA’s in a cell. This provides functional information and therapeutic insights for therapy. The tool uses mutant RISC complex allowing microRNA to just bind to its target but no further processing takes place [133].

<table>
<thead>
<tr>
<th>Step</th>
<th>Tool</th>
<th>Function</th>
<th>Algorithm/ Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Reads</td>
<td>NGSQC[134]</td>
<td>Analysis of sequence quality</td>
<td>Pearl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GC content</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presence of adaptors</td>
<td></td>
</tr>
<tr>
<td>Read alignment</td>
<td>Picard[135]</td>
<td>Quality control in mapping</td>
<td>Java</td>
</tr>
<tr>
<td></td>
<td>RSeQC[136]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Qualimap[137]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantification</td>
<td>NOISEq[134]</td>
<td>Quality control of count data.</td>
<td>SEECER error correction algorithm</td>
</tr>
<tr>
<td></td>
<td>EDASeq[138]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genome Mapping</td>
<td>TopHat[123]</td>
<td></td>
<td>TopHat Fusion Algorithm</td>
</tr>
<tr>
<td></td>
<td>STAR[139]</td>
<td></td>
<td>Sequential maximum mappable seed search in uncompressed suffix arrays followed by seed clustering and stitching procedure Algorithm</td>
</tr>
<tr>
<td></td>
<td>BowTie[140]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcript discovery</td>
<td>GRT[144]</td>
<td>Identify transcription start and end sites</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td></td>
<td>Cufflinks[145]</td>
<td></td>
<td>Cufflink Algorithm</td>
</tr>
<tr>
<td></td>
<td>StringTie[146]</td>
<td></td>
<td>Network flow Algorithm and de novo assembly step Likelihood based Monte Carlo Algorithm</td>
</tr>
<tr>
<td></td>
<td>Montebello[147]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4- Discussion of tools and their function, along with the step at which they are used and algorithm applied.

4.1 Pipelines associated with Next Generation Sequencing

1) Mutation T@ster-This software handles sequences in FASTA, FASTQ and CSFASTA format i.e., ABI SOLiD, Roche 454 and Illumina Genome Analyzer reads. There are more efficient alignment algorithms available than BLAST, but most of them are platform-specific and fails to handle arbitrary length reads. For speeding up the alignment procedure, the reference sequence is reduced to a set of target regions and additionally reads are split into smaller chunks to parallelize the alignment step.

2) NARWHAL-This pipeline has been developed for automation of primary analysis of data obtained through Illumina sequencing. This pipeline combines a novel and flexible de-multiplexing tool with open-source aligners and provides automated quality assessment. The entire pipeline can be run using only one simple sample-sheet for diverse sequencing applications. NARWHAL creates a sample-oriented data structure and outperforms existing tools in speed. [155].

3) DSAP(deep-sequencing small RNA analysis pipeline)-This is an automated multiple-task web service designed to provide a total solution to analyzing deep-sequencing small RNA datasets generated by NGS technology. The input form is in tab-delimited file, which holds the unique sequence reads generated by the Solexa sequencing platform, and their corresponding thier number of copies. [156]
4) DDBJ Read Annotation Pipeline-The public NGS reads of the DDBJ Sequence Read Archive located on the same supercomputer can be imported by inputting the accession number only. The research will be facilitated by using this proposed pipeline, by utilizing unified analytical workflows applied to the NGS data. [157]

V. CONCLUSION:

The humongous genetic data generated by Next Generation Sequencing has made a noteworthy impression on the clinical diagnoses while parallelly contributing to the discovery of molecular pathomechanisms fundamental to these diseases. In present times, Neurodegenerative diseases are increasing in their prevalence but their treatment options are still limited. With the advent of Next Generation Sequencing, there is a hope for improvement in the current techniques in order to target specific genes related to the disease, which may be under or over expressed. Study of mutations in particular genes related to various diseases like Alzheimer’s (FAD, PSEN1, and PSEN), Parkinson’s (PINK1-PARK6), Ataxia (FXM, ATM) and Multiple Sclerosis (IL7R, IL2R, CD226, TKY2) helps to broaden our perspective towards these least curative diseases. Being very cost effective and time saving Next Generation Sequencing have revolutionised the study of Genomics and Molecular Biology. NGS technologies function towards generation of short sequences with higher error rates. Even though the price of the instrument is low, but the overall investment for the sequencing of single genome is very high. Costs per base for sequencing are generally higher than the standard instrument and hence an overall infrastructure is still required [158]. The current technology poses inherent challenges especially in handling enormous amounts of data. However, massively parallel sequencing platforms and development in technology, have facilitated with novel hope to a certain extent. Currently, more advancement is desirable to understand and interpret the data correctly.

ACKNOWLEDGEMENT

We are thankful to Dr. Chakresh Kumar Jain for mentoring us and guiding us throughout the paper. We sincerely thank Department of Biotechnology, Jaypee Institute of Information Technology, NOIDA, for providing the conducive environment

REFERENCES

[16] Bennett, S.T., Barnes, C., et al., 2005, Toward the 1,000 dollars human genome, Pharmacogenomics, Vol 6:pp373–382


Leidinger, P., Backes, C., et al., 2013, A blood based 12-miRNA signature of Alzheimer disease patients, Genome Biology, Vol 14[R78]:pp1-16


[71] Symptomatic pharmacological therapy in Parkinson’s disease, Parkinson's Disease London: Royal College of Physicians. pp39–100
[100] Ritchie, D.M., 2009, Using prior knowledge and genome wide association to identify pathways involved in multiple sclerosis, Genome Medicine, Vol 1:pp63
inversion recovery MR imaging at 3 Tesla, European Radiology, Vol 20:pp1675-1680


[114] www.alz.co.uk/research/statistics


[120] Capurro, A., et al., 2015, Computational deconvolution of genome wide expression data from Parkinson's and Huntington's disease brain tissues using population-specific expression analysis, Frontiers in Neuroscience, Vol 8; pp-441


