Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

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, MRE-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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

80-120 Mb sequences in 200 to 300 reads in a 4 hour run. Illumina/Solexa approach [16] can amplify DNA without cloning and by attachment of single stranded DNA fragments to a solid surface. The current achievable read length is approximately 300bp [17].Though Illumina gives better results at homopolymeric fragment level sequencing than pyrosequencing, it produces shorter read lengths [18]. Since NGS has provided a cheaper high throughput, its application in many research areas has expanded. cDNA can also be sequenced rather than the whole genome (which reduces the target sequencing size). Paired end approaches are also used where a definite size is sequenced at both ends to give more information about the fragment, and can significantly improve the usage of shorter reads for de novo sequence assembly and rearranged genomic segments [19]**.** Table 1 below compares various NGS platforms:

Technology	Principle	Analysis time	Average length read	Throughput (Mb/hr)
Roche titanium	Pyrosequencing	4 hours	300-500bp	20-30
Illumina MiSeq	Bridge amplification	27hours	Upto 150bp	20
Ion Torrent	Incorporation of H^+ ion	2 hours	200bp	320

Table 1- Comparison between various NGS platforms [15]

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.

Condition	Result
Protein misfolding and/or defective degradation	Disruption of cellular/axonal transport
Oxidative stress and formation of free radicals	Actions and mutations of molecular chaperones
Mitochondrial dysfunctions	Dysfunction of neurotrophins
Fragmentation of neuronal Golgi apparatus	Neuro-immune processes Source

Table 2- Etiology of neurodegenerative disease [27]

1. *Alzheimer's disease*

Alzheimer's is a widespread functional disease of the human nervous system. There are alot 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].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

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 (Aß), formed by sequential amyloid precursor protein proteolysis and hyperphosphorylated Tau proteins [42]. Use of markers like Aß 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, miRNAs are generally protected from degradation due to their binding with RNA binding proteins, e.g. Lipoproteins [51]. In AD, miRNAs 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 miRNA 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 performed 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].

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

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]

Various tools and scripts such as Perl are available for processing huge amounts of data by a streamlined work-flow. [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.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

Primary Parkinson may be genetic or idiopathic while secondary Parkinson is mainly caused due to toxins. Accumulation of levy bodies in neurons of the mid-brain hampers dopamine-production causing the disease. Causes can be demarcated clinically and etiologically, including idiopathic paralysis agitan, 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, occulo-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). Cryothalamectomy 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]. Friedreich 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.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

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 with 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.

Disease	Genes	Symptoms	People affeted
Alzheimer's	FAD, apolipoprotein E (APOE), two presenilin (PSEN-1 and PSEN-2)	Difficulty to retain information, apathy and exhaustion, sun downing.	46.8 million by 2015
Parkinson's	PINK1 (PARK6)	Resting tremor, Bradykinesia, Postural instability, Micrographic.	7-10 million by 2015
Ataxia	FXN, ATM, Abcb7	Impaired limb coordination, Cognitive and Mood Problems, Fine motor incardination.	-
Multiple Sclerosis	IL7R, IL2R, CLEC1 6A, CD226 , TKY2	Tingling, numbness, blurred vision, muscle stiffness.	2.5 million by 2015

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.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

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 corrected 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 tarageted genes and the associated miRNAs. Tools such as CORNA [129], MMIA[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].

Step	Tool	Function	Algorithm/ Program
Raw Reads	NGSQC[134]	Analysis of sequence quality GC content Presence of adaptors	Pearl
Read alignment	Picard[135] RSeQC[136] Qualimap[137]	Quality control in mapping	Java
Quantification	NOISeq[134] EDASeq[138]	Quality control of count data.	SEECER error correction algorithm
Genome Mapping	TopHat[123] STAR[139] BowTie[140]		TopHat Fusion Algorithm Sequential maximum mappable seed search in uncompressed suffix arrays followed by seed clustering and stitching procedure Algorithm
Transcript discovery	GRIT[144] Cufflinks[145] StringTie[146] Montebello[147]	Identify transcription start and end sites	Genetic Algorithm Cuflink Algorithm Network flow Algorithm and de novo assembly step Likelihood based Monte Carlo Algorithm
De novo transcription	SOAPdenovo-Trans [148] Oases[149] Trans-ABySS[150] Trinity[151]	RNA-seq reads can be assembled de novo	Short Read Assemblers and DE-Bruijin Graph
Alternative splicing analysis	DEXseq DSGSeq[152]	Transcrip-level differential expression analysis	Negative bionomial distribution
Differential gene expression analysis	COMBAT[141] ARSyN[142][143]	Comparing and normalising gene Expression values.	COMBAT Algorithm using parametric and non-parametric empirical Bayes framework Eigentaste Algorithm
Functional profiling with RNA-seq	Gene Set Variation Analysis (GSVA) [153] SeqGSEA[154]	Characterization of the molecular functions or pathways in which differentially expressed genes	GSVA Algorithm Negative bionomial distribution

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]

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

4) DDBJ Read Annotation Pipeline-The public NGS reads of the DDBJ Sequence Read Archive located on the same supercomputer can be imported by inputing 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 parallely 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

[1] Ries-Filho, S.J., 2009, Next-generation sequencing, Breast Cancer Research, Vol 11:S12
[2] Schuster, C.S., 2008, Next Generation Sequencing transforms today's Biology, Nature, Vol 5:pp16-18
[3] Sanger, F., Nicklen, S., et al., 1997, DNA Sequencing with chain-termination inhibitors, Proceedings of National Academy of Sciences of the United States of America, Vol 74:pp5436-67
[4] Mardis, R.E., 2008, Impact of Next Generation Sequencing on genetics, Cell Symposia, Vol 24:pp133-141
[5] Quail, A.M., Smith, M., et al., 2012, A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers, BioMed Central Genomics, Vol 13:pp341
[6] Morozova, O., Marra, M.A., 2008, Applications of next-generation sequencing technologies in functional genomics, Elsevier Genomics, Vol 92:pp255-264
[7] Hall, N., 2000, Advanced sequencing technologies and their wider impact in microbiology, Journal of Experimental Biology, Vol 210:pp1518–1525
[8] Harris, R.A., Wang, T., et al., 2010, Comparison of sequencing based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nature Biotechnolgy, Vol 28:pp 1097–1105
[9] Fouse, S.D., Nagarajan, R.O., et al., 2010 , Genome-scale DNA methylation analysis, Epigenomics, Vol 2:pp 105–117
[10] Barski, A., Cuddapah, S., et al. ,2007,High-resolution profiling of histone methylations in the human genome, Cell, Vol 129: pp 823-837
[11] Margulies, M., Egholm, M., et al., 2005, Genome sequencing in microfabricated high-density picolitre reactors, Nature, Vol 437:pp376–380
[12] Tawfik, D.S., Griffiths, A.D., 1989, Man-made cell-like compartments for molecular evolution, Nature Biotechnology, Vol 16:pp652–656
[13] Nyren, P., Pettersson, B., Uhlen, M., 1993, Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay, Elsevier, Vol 208:pp171–175
[14] Ronaghi, M., et al., 1996, Real-time DNA sequencing using detection of pyrophosphate release, Elsevier, Vol 242:pp84–89
[15] bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-341
[16] Bennett, S.T., Barnes, C., et al., 2005, Toward the 1,000 dollars human genome, Pharmacogenomics, Vol 6:pp373–382
[17] Wold, B., Myers, R.M., 2008, Sequence census methods for functional genomics, Nature, Vol 5:pp19–21
[18] Bentley, D.R., 2006, Whole-genome re-sequencing, Elsevier, Vol 16:pp545–552
[19] Hall, N., Mardis, E.R., 2006, Anticipating the 1,000 dollar genome, Genome Biology, Vol 7:pp112
[20] Uttara, B., Singh, A.V., et al., 2009, Oxidative Stress and Neurodegenerative Diseases: A Review of Upstream and Downstream Antioxidant Therapeutic Options, Current Neuropharmacology, Vol 7:pp65-74
[21] Daar, A.S., Singer,P.A., et al., 2007, Grand challenges in chronic non-communicable diseases, Nature, Vol 450:pp494-496
[22] Levenson, R.W., Sturm, V.E., Haase, C.M., 2014, Emotional and behavioural symptoms in neurodegenerative disease: a model for studying the neural bases of psychopathology, Annual Review of Clinical Psychology, Vol 10:pp581-606
[23] Hara, T., Nakamua, K., et al., 2006, Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice, Nature, Vol 441:pp885-889
[24] Chin, J.H., 2004, The Growing Global Burden of Neurologic Disease, Neurology, Vol 87:pp453
[25] Brown, R., et al., 2005, Neurodegenerative Diseases:An Overview of Environmental Risk Factors, Environmental Health Perspectives, Vol 113:pp1250-1256
[26] Behari, M., et al., 2001, Risk factors of Parkinson's disease in Indian patients, Elsevier, Vol 190:pp49-55
[27] Bernal, G.M., Peterson, D.A., 2004, Neural stem cells as therapeutic agents for age-related brain repair. Aging Cell ,Vol 3(6):pp 345-351
[28] Brookmeyer, R., et al., 2007, Forecasting the global burden of Alzheimer's disease, Elsevier, Vol 3:pp186-191 .
[29] Querfurth, H.W., LaFerla, F.M., 2010, Alzheimer's disease, New England Journal of Medicine, Vol 362:pp329-344

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

[30] Herrup, K., 2010, Reimagining Alzheimer's disease-An age based Hypothesis, The journal of Neuroscience, Vol 30:pp16755-62
[31] Cataldo, J.K., Prochaska, J.J., Glantz, S.A., 2010, Cigarette smoking is a risk factor for Alzheimer's Disease: an analysis controlling for tobacco industry affiliation, Journal of Alzheimers Disease, Vol 19:pp465-480
[32] Lindsay, J., Laurin, D., et al., 2002, Risk factors for Alzheimer's disease: A prospective analysis from the Canadian Study of Health and Aging, American Journal of Epidemiology, Vol 156:pp445-453
[33] Podewils, L.J., Guallar, E., et al., 2005, Physical activity, APOE genotype, and dementia risk: Findings from the Cardiovascular Health Cognition Study, American Journal of Epidemiology, Vol 161:pp639-651
[34] Scarmeas, N., Stern, Y., et al., 2006, Mediterranean diet, Alzheimer disease, and vascular mediation, Arch Neurology, Vol 63:pp1709-1717
[35] Patterson, C., Fegihtner, J., et al., 2007, General risk factors for dementia: a systematic evidence review, Elsevier, Vol 3:pp341-347
[36] Villemagne,V., Bumham, S., et al., 2013, Amyloid β deposition neuro degeneration and cognitive decline in sporadic Alzheimer's disease: A prospective cohort study, The Lancet Neurology, Vol 12:pp357–367
[37] Brookmeyer, R., Gray, S., Kawas, C., 1998, Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset, American Journal of Public Health, Vol 88:pp1337-1342
[38] Simón-Sánchez, J., Singleton, A., 2008, Genome-wide association studies in neurological diseases, Lancet Neurology, Vol 7[1]:pp1067-72
[39] Jiang, T., Tan, S.M., et al., 2014, Application of next-generation sequencing technologies in Neurology, Annals of Translational Medicine, Vol 2[12]:pp125-136
[40] Metzker, M.L., 2010, Sequencing technologies - The Next Generation, Nature Review Genetics, Vol 11[31]:pp31-46
[41] Hollingworth, P., Harold, D., et al., 2011, Alzheimer's disease genetics: current knowledge and future challenges, International Journal of Geriatric Psychiatry, Vol 26:pp793-802
[42] Lupton, M.K., Proitsi, P., et al., 2011, Deep sequencing of the Nicastrin gene in pooled DNA, the identification of genetic variants that affect risk of Alzheimer's disease, PLoS One, Vol 6:e17298(2011).
[43] Hardy, J., Allsop, D., 1991, Amyloid deposition as the central event in the aetiology of Alzheimer's disease, Trends in Pharmacological Sciences, Vol 12:pp383–388
[44] Malaplate-Armand, C., et al., 2009, Biomarkers for early diagnosis of Alzheimer's disease:Current update and future direction, Revue Neurologique, Vol 165:pp511-520
[45] Gasparini, L., Racchi, M., et al., 1998, Peripheral markers in testing pathophysiological hypotheses and diagnosing Alzheimer's disease, Federation of American Societies for Experimental Biology, Vol 12:pp17-34
[46] Leidinger, P., Backes, C., et al., 2013, A blood based 12-miRNA signature of Alzheimer disease patients, Genome Biology, Vol 14[R78]:pp1-16
[47] Skog, J., Wurdinger, T., et al., 2008, Glio blastoma micro vesicles transport RNA and proteins that promotet tumour growth and provide diagnostic biomarkers, Nature Cell Biology, Vol 10
[48] He, L., Hannon, G.J., 2004, Micro RNAs:Small RNAs with a big role in gene regulation, Nature Reviews Genetics, Vol 5:pp522–531
[49] Cheng, L., Quek, C.Y., et al., 2013, The detection of microRNA associated with Alzheimer's disease in biological fluids using next-generation sequencing technologies, Frontiers in Genetics, Vol 4[150]:pp1-11
[50] Babak, T., Zhang, W., et al., 2004, Probing microRNAs with micro arrays, RNA, Vol 10:pp1813–19
[51] Makeye, E.V., Zhang, J., et al., 2007, The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing, Molecular Cell, Vol 27:pp435–448
[52] Arroyo, J.D., Chewillet, J,R., et al., 2011, Argonaute 2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma, Proceedings of the National Academy of Sciences of the USA, Vol 108:pp5003–5008
[53] Blennow, K., Wallin, A., et al., 1990, Blood-brain barrier disturbance in patients with Alzheimer's disease is related to vascular factors, ActaNeurologica Scandinivica, Vol 81:pp323–326
[54] Haqqani, A.S., Delaney, C.E., et al., 2013, Method for isolation and molecular characterization of extracellular microvesicles released from brain endothelial cells, Fluids and Barriers of CNS, Vol 10[4]
[55] Schipper, H.M., Maes, O.C., et al., 2007, MicroRNA expression in Alzheimer blood mononuclear cells, Gene Regulation and Systems Biololgy, Vol 1:pp263–274
[56] Geekiyanage, H., Jichas, G.A., et al., 2012, Blood serum miRNA: Non-invasive biomarkers for Alzheimer's disease, Experimental Neurology, Vol 235:pp491–496
[57] Cogswell, J.P., Ward, J., et al., 2008, Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways, Journal of Alzheimers Disease, Vol-14:pp27–41
[58] Gallego, J., Gordan, M.L., et al., 2012, In vivo microRNA detection and quantitation in cerebrospinal fluid, Journal of Molecular Neuroscience, Vol 47:pp243–248
[59] Lehmann, S., Kruger, C., et al., 2012, An unconventional role for miRNA: let-7activates Toll-like receptor 7 and causes neurodegeneration, Nature Neuroscience, Vol 15: pp827–835
[60] Mitchel, P., Parkin, R., et al., 2008, Circulating microRNAs as stable blood-based markers for cancer detection, Proceedings of the National Academy of Sciences of the USA, Vol 105:pp10513–10518
[61] Zhang, J., Chiodini, R., et al., 2014, The impact of next-generation sequencing on genomics, Journal of Genetic Genomics, Vol 38
[62] Pearce, J., 1977, Symptomatic Parkinsonism, Postgraduate Medical Journal, Vol 53:pp726-727.
[63] Charles, H., et al, 1974 Parkinson's Disease and Levodopa. Medical Progress. Vol-121:pp188-206.
[64] Hughes, AJ., Daniel, S., et al., 2002, The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disease service, Brain, Vol 125:pp861–70
[65] Ambros V., 1989, A hierarchy of regulatory genes controls a larva to adult development switch in C.elegans, Cell, Vol 57:pp49-57
[66] Davie, CA., 2008, A review of Parkinson's disease, British medical Bulletin, Vol 86:pp109–27
[67] Lesage, S., Brice, A., 2009, Parkinson's disease: from monogenic forms to genetic susceptibility factors, Human Molecular Genetics, Vol 18:pp48–59
[68] Grozdanov, V., Bliederhaeuser, C., et al., 2014, Inflammatory dysregulation of blood monocytes in Parkinson's disease patients, Acta Neuropathologica, Vol 128:pp651-663

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

[69] Weismer, G., Jeng, J.Y., et al., 2001, Acoustic and intelligibility Characteristics of Sentence Production in Neurogenic Speech Diseases, Folia Phoniatrica Lopoget Logopaedica, Vol 53:pp1-18
[70] Poewe, W., 2006, The natural history of Parkinson's disease, Journal of Neurology, Vol253
[71] Symptomatic pharmacological therapy in Parkinson's disease, Parkinson's Disease London: Royal College of Physicians. pp59–100
[72] Zhiming, L., Qing, L., et al., 2015, Target Gene Capture Sequencing in Chinese Population of Sporadic Parkinson Disease, Medicine, Vol 94:pp1-6
[73] McFarlin D.E., Strober W., et al., 1972, Medicine. Vol 51:pp281–314 .
[74] Taylor, A.M.R., Harnden, D., et al., 1975, Ataxia telangiectasia: a human mutation with abnormal radiation sensitivity, Nature, Vol 258:pp427 - 429
[75] Anheim, M., Fleury, M., et al., 2010, Epidemiological, clinical, paraclinical and molecular study of a cohort of 102 patients affected with autosomal recessive progressive cerebellar ataxia from Alsace, eastern France: implications for clinical management, Neurogenetics, Vol 11:pp1–12
[76] Silva, M.C., Coutinho, P., et al., 1997, Hereditary ataxias and spastic paraplegias: Methodological aspects of a prevalence study in Portugal, Journal of Clinical Epidemiology, Vol 50:pp1377–84
[77] Sequeiros, J., Martins, S., et al., 2012, Epidemiology and population genetics of degenerative ataxias, Handbook of Clinical Neurology, Vol 103:pp227–51
[78] Anheim, M., Fleury, M., et al., 2010, Epidemiological, clinical, paraclinical and molecular study of a cohort of 102 patients affected with autosomal recessive progressive cerebellar ataxia from Alsace, eastern France: implications for clinical management, Neurogenetics, Vol 11:pp1–12
[79] Himmelmann, K., Hagberg, G., Uvebrant, P., 2010, The changing panorama of cerebral palsy in Sweden. X. Prevalence and origin in the birth-year period 1999–2002, Acta Pediatrica, Vol 99:pp1337–43
[80] Esscher, E., Flodmark, O., et al.,1996, Non-progressive ataxia: origins, brain pathology and impairments in 78 swedish children, Developmental Medicine and Child Neurology, Vol 38:pp285–96
[81] Fogel, B.L., Geschwind, D.H., 2012, Clinical neurogenetics, Neurology in Clinical Practice. Vol-6:pp704–734(2012).
[82] Coppola, G., Geschwind, D.H., 2012, Genomic medicine enters the neurology clinic, Neurology, Vol 79(2):pp112–114
[83] Fogel, B.L., Vickrey, B.G., et al., 2013, Utilization of genetic testing prior to subspecialist referral for cerebellar ataxia, Genetic Testing and Molecular Biomarkers, Vol 17(8):pp588–594
[84] Sailer, A., Scholz, S.W., et al., 2012, Exome sequencing in an SCA14 family demonstrates its utility in diagnosing heterogeneous diseases, Neurology, Vol 79(2):pp127–131
[85] Chen, Z., Wang, J.L., et al., 2013, Using next-generation sequencing as a genetic diagnostic tool in rare autosomal recessive neurologic Mendelian diseases, Neurobiology of Aging, Vol 34(10):pp2442.e11–7
[86] Fogel, B.L., Lee, H., et al., 2014, Clinical exome sequencing: the new standard in genetic diagnosis, Annals of the NewYork academy of Sciences, Vol 71:pp1237-46
[87] Galindo, B.E., Vacquier, V.D., 2005, Phylogeny of the TMEM16 protein family: some members are overexpressed in cancer, International Journal of Molecular Medicine, Vol 16:pp919–924
[88] Hartzell, H.C., Yu, K., et al., 2009, Anoctamin/TMEM16 family members are Ca2+ activated Cl- channels. Journal of Physiology, Vol 587:pp2127–2139
[89] Vermeer, S., Hoischen, A., et al., 2010, Targeted Next-Generation Sequencing of a 12.5 Mb Homozygous Region Reveals ANO10 Mutations in Patients with Autosomal-Recessive Cerebellar Ataxia, American Journal of Human Genetics, Vol 87(6):pp813–819
[90] Moreira, M.C., Barbot, C., et al., 2001, The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin, Nature Genetics, Vol 29(2):pp189-93
[91] Waters, M.F., Minassian, N.A., et al., 2006, Mutations in voltage-gated potassium channel KCNC3 cause degenerative and developmental central nervous system phenotypes, Nature Genetics, Vol 38(4):pp447-51
[92] Figueroa, K.P., Minassian, N.A., et al., 2010, KCNC3: phenotype, mutations, channel biophysics-a study of 260 familial ataxia patients, Human Mutation, Vol-31(2):pp191-96
[93] Nemeth, A.H., Kwasniewska, A.C., et al., 2013, Next generation sequencing for molecular diagnosis of neurological diseases using ataxias as a model, Brain, Vol 136:pp3106-3118
[94] Jacquemin, V., Rieunier, G., et al., 2012, Underexpression and abnormal localization of ATM products in ataxia telangiectasia patients bearing ATM missense mutations, European Journal of Human Genetics, Vol 20(3):pp305-12
[95] Collins, F.S., Varmus, H., 2015, A new initiative on precision medicine, The New England Journal of Medicine, Vol-372(9):pp793-5
[96] Chen, Z., Ye, W., et al., 2015, Targeted Next-Generation Sequencing Revealed Novel Mutations in Chinese Ataxia Telangiectasia Patients: A Precision Medicine Perspective, PLoS ONE, Vol 10,e01397389(2015)
[97] Rosenberg R.N., Ataxic diseases. Harrison's Principles of Internal Medicine. e14:pp2365-2367(1998).
[98] Filla, A., Michele, G., et al., 1996, The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia, American Journal of Human Genetics, Vol 59:pp554 -560
[99] Fischer, S.J., Rudick, R.A., et al., 2000, The Multiple Sclerosis Functional Composite measure (MSFC): an integrated approach to MS clinical outcome assessment, Multiple Sclerosis, Vol 5:pp 244-250
[100] Ritchie, D.M., 2009, Using prior knowledge and genome wide association to identify pathways involved in multiple sclerosis, Genome Medicine, Vol 1:pp65
[101] Mero, I.L., Lorentzen, A.R., et al., 2010, A rare variant of the TYK2 gene is confirmed to be associated with multiple sclerosis, European Journal of Human Genetics, Vol 18:pp502-4
[102] Kamb, A., 2013, Next-Generation Sequencing and Its Potential Impact, Chemical Research in Toxicology, Vol 16:pp2119-27
[103] Barcellos, L.F., Oksenberg, J.R., et al., 2003, HLA-DR2 Dose Effect on Susceptibility to Multiple Sclerosis and Influence on Disease Course, American Journal of Human Genetics, Vol 72:pp-716-719
[104] Zivadinov, R., Bakshi, R., 2004, Role of MRI in multiple sclerosis I: inflammation and lesions, Frontiers of Bioscience, Vol 9:pp665-683
[105] Calabrese, M., Gallo, P., 2009, Magnetic resonance evidence of cortical onset of multiple sclerosis, Multiple Sclerosis, Vol 15:pp933-941
[106] Bo, L., Vedeler, A.C., et al., 2003, Intracortical multiple sclerosis lesions are not associated with increased lymphocyte infiltration, Multiple Sclerosis, Vol 9:pp323-331
[107] Kutzelnigg, A., Lucchinetti, C.F., et al., 2005, Cortical demyelination and diffuse white matter injury in multiple sclerosis, Brain, Vol 128:pp2705-2712
[108] Simon, B., Schmidt, S., et al., 2010, Improved in vivo detection of cortical lesions in multiple sclerosis using double

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

inversion recovery MR imaging at 3 Tesla, European Radiology, Vol 20:pp1675-1680
[109] Tallantyre, C.E., Morgan, P.S., et al., 2010, 3 Tesla and 7 Tesla MRI of Multiple Sclerosis Cortical Lesions, Journal of magnetic Resonance Imaging, Vol 32:pp971–977
[110] Ota, K., Matsui, M., et al., 1990, T-cell recognition of an immune dominant myelin basic protein epitope in multiple sclerosis, Nature, Vol 346:pp183-187
[111] Steinman, L., 2001, Multiple Sclerosis: a two-staged disease, Nature Immunology, Vol 2:pp762-764
[112] Amato, M.P., Ponziani, G., et al., 2001, Quality of life in Multiple Sclerosis:the effect of depression, fatigue and disability, Multiple Sclerosis, Vol 7:pp340-344
[113] Kurtzke, J.F., 1983, Rating neurologic impairment in multiple sclerosis, Neurology, Vol-33
[114] www.alz.co.uk/research/statistics
[115] www.pdf.org/en/parkinson_statistics
[116] wwww.healthline.com/health/multiple-sclerosis/facts-statistics-infographic
[117] Z. Wang, M. Gerstein and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics", Nat Rev Genet, vol. 10, no. 1, pp. 57-63, 2009.
[118] Andrews, S., FASTQC. A quality control tool for high throughput sequence data. http://www.bioinformatics. babraham.ac.uk/projects/fastqc/. Accessed 29 September 2014.
[119] Dai, M., Thompson, R.C., et al., 2010, NGSQC: cross-platform quality analysis pipeline for deep sequencing data, BMC Genomics, Vol 11:pp4-7
[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
[121] Roberts, A., Pimentel, H., et al., 2011, Identification of novel transcripts in annotated genomes using RNA-Seq, Bioinformatics, Vol 27:pp2325–9.
[122] Trapnell, C., Williams, B.A., et al., 2010, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nature Biotechnology, Vol 28:pp511–5.
[123] Backes, C., et al., 2015, miFRame:analysis and visualization of miRNA sequencing data in neurological disorders, Journal of Translational Medicine,Vol 13:pp-224
[124] Li, B., Dewey, C.N., 2011, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinformatics, Vol12:pp323.
[125] Roberts, A., Pachter, L., 2013, Streaming fragment assignment for real-time analysis of sequencing experiments, Nature Methods, Vol 10:pp71–3.
[126] Anders, S., Pyl, P.T., 2015, HTSeq - a Python framework to work with high-throughput sequencing data, Bioinformatics, Vol 31:pp166–9.
[127] Bray N, Pimentel H, Melsted P, Pachter L. Near-optimal RNA-Seq quantification with kallisto. https://liorpachter. wordpress.com/2015/05/10/near-optimal-rna-seq-quantification-with-kallisto/. Accessed 6 January 2016.
[128] Ma, X., Zhang, X., 2013, NURD: an implementation of a new method to estimate isoform expression from non-uniform RNA-seq data, BMC Bioinformatics, Vol 14:pp220.
[129] Wu, X., Watson, M., 2009, CORNA: testing gene lists for regulation by microRNAs, Bioinformatics, Vol 25:pp832–3.
[130] Lee, H., Yang, Y., Chae, H., et al., 2012, BioVLAB-MMIA: a cloud environment for microRNA and mRNA integrated analysis (MMIA) on Amazon EC2, IEEE Trans Nanobioscience, Vol 11:pp266–72.
[131] Nam, S., Li, M., et al., 2009, MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression, Nucleic Acids Research, Vol 37:ppW356–62.
[132] Icay, K., Chen, P., et al., 2015, SePIA: RNA and smallRNA-sequence processing, integration, and analysis.
[133] Hendrix, D., Levine, M., Shi, W., 2010, miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data, Genome biology, Vol 11: pp-4.
[134] Tarazona, S., Furió-Tarí, P., et al., 2015, Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package, Nucleic Acids Research, Vol 43:pp140.
[135] Picard. http://picard.sourceforge.net/. Accessed 12 January 2016.
[136] Wang, L., Wang, S., et al., 2012, RSeQC: quality control of RNA-seq experiments, Bioinformatics, Vol 28:2184–5.
[137] García-Alcalde, F., Okonechnikov, K., et al, 2012, Qualimap: evaluating next-generation sequencing alignment data, Bioinformatics, Vol 28:pp2678–9.
[138] Risso, D., Schwartz, K., et al., 2011, 0GC-content normalization for RNA-seq data, BMC Bioinformatics, Vol 12:pp480.
[139] Dobin, A., Davis, C.A., et al., 2013, STAR: ultrafast universal RNA-seq aligner, Bioinformatics, Vol 29:pp15–21.
[140] Langmead, B., Salzberg, S.L., 2012, Fast gapped-read alignment with Bowtie, Nature Methods, Vol 9:pp357–9.
[141] Johnson, W.E., Rabinovic, A., et al., 2007, Adjusting batch effects in microarray expression data using Empirical Bayes methods, Biostatistics, Vol 8:pp118–27.
[142] Risso, D., Schwartz, K., Sherlock, G., et al., 2011, GC-content normalization for RNA-seq data, BMC Bioinformatics, Vol 12:480.
[143] Nueda, M.J., Ferrer, A., et al., 2012, ARSyN: a method for the identification and removal of systematic noise in multifactorial time course microarray experiments, Biostatistics, Vol 13:pp553–66.
[144] Boley, N., Stoiber, M.H., et al., 2014, Genome-guided transcript assembly by integrative analysis of RNA sequence data, Nature Biotechnology, Vol 32:pp341–6.
[145] Roberts, A., Pimentel, H., et al., 2011, Identification of novel transcripts in annotated genomes using RNA-Seq, Bioinformatics, Vol 27:2325–9.
[146] Pertea, M., Pertea, G.M., et al., 2015, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nature Biotechnology, Vol 33:290–5.
[147] Hiller, D., Wong, W.H., 2013, Simultaneous isoform discovery and quantification from RNA-Seq, Statistical Biosciences, Vol 5:pp100–18.
[148] Xie, Y., Wu, G., et al., 2014, SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads, Bioinformatics, Vol 30:pp1660–6.
[149] Schulz, M.H., Zerbino, D.R., et al., 2012, Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels, Bioinformatics, Vol 28:pp1086–92.
[150] Grabherr, M.G., Haas, B.J., et al., 2011, Full-length transcriptome assembly from RNA-seq data without a reference genome, Nature Biotechnology, Vol 29:pp644–52.
[151] Haas, B.J., Papanicolaou, A., et al., 2013, De novotranscript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, Nature Protocols, Vol 8:pp1494–512.
[152] Trapnell, C., Roberts, A., et al., 2012, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, Nature Protocols, Vol 7:pp562–78.
[153] Hänzelmann, S., Castelo, R., et al., 2013, GSVA: gene set variation analysis for microarray and RNA-Seq data, BMC Bioinformatics, Vol 14:pp7.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

[154] Wang, X., Cairns, M.J., 2013, Gene set enrichment analysis of RNA-Seq data: integrating differential expression and splicing, BMC Bioinformatics, Vol 14:ppS16.
[155] Brouwer, R.W.W., Van, N., et al., 2011, NARWHAL:a primary analysis pipeline for NGS data, Oxford Journal.
[156] Huang, P.J., Liu, Y.C., et al., 2010, DSAP: deep-sequencing small RNA analysis pipeline, Oxford Journal.
[157] Nagasaki, H., Mochizuki, T., et al., 2013, DDBJ Read Annotation Pipeline:A Cloud Computing-Based Pipeline for High-Throughput Analysis of Next-Generation Sequencing Data, Oxford Journal.
[158] Kircher, M., Kelso, J., 2010, High Throughput DNA sequencing- concepts and limitations, Bioessays, Vol 32: pp 524-536.