Roger Torra

Identification of methylation differences

Abstract

Neurodegenerative diseases arise from a complex interplay of genetic, environmental, and epigenetic factors. Epigenetic mechanisms, such as DNA methylation, provide a link between the genome and external influences, offering insights into disease mechanisms. This study focuses on methylation patterns in PRKAR1A, a gene involved in key regulatory pathways, and chromosome 17, across four neurodegenerative diseases: Down syndrome, Alzheimer's disease, Dementia of Lewy bodies, and Parkinson's disease.

Illumina HiSeq 2000 sequencing data were processed to align reads to the GRCh37 (hg19) reference genome. The results showed that Parkinson's disease exhibited the highest chromosome-wide methylation (10.45%), while Alzheimer's and Dementia of Lewy bodies demonstrated elevated methylation in PRKAR1A. However, due to the lack of replicates and a control group, statistical significance could not be assessed.

This study highlights potential epigenetic links between methylation in PRKAR1A and neurodegenerative diseases but emphasizes the need for future research with biological replicates and healthy controls to validate these findings.

Motivation

Neurodegenerative diseases are disorders caused by a convergence of several factors, such as aging, genetic and environmental factors. Normally these factors individually do not generate neurodegenerative diseases, however the combination of some of them may lead to the onset of the disease. Epigenetics, the study of gene expression depending on external factors, acts as a mediator between the genome and the environment. It provides a mechanistic explanation that offers a distinct point of view of the situation and allows us to understand the disease more. By looking into epigenetically deregulated genes emerging on neurodegenerative diseases, we can discover their differences and similarities.

The current object of the study is to Identify Differentially Methylated Sites between four diseases: Down's syndrome, Parkinson's disease, Dementia of Lewy bodies and Alzheimer's disease. Previous studies have revealed that an altered cyclic nucleotide signaling pathway is related to neurodegenerative diseases. This pathway is illustrated by the kinase cAMP-dependent regulatory type I alpha (PRKAR1A) and the cGMP-dependent protein kinase type II (PRKG2) genes.

The PRKAR1A gene produces a signalling molecule key for a variety of functions. This protein was found to be a tissue-specific extinguisher that down-regulates the expression of seven liver genes in hepatoma x fibroblast hybrids. Mutations in this gene cause the Carney

complex (CNC). A nonconventional nuclear localization sequence (NLS) has been found for this protein which suggests a role in DNA replication via the protein serving as a nuclear transport protein for the second subunit of the Replication Factor C (RFC40). Additionally, this gene can fuse to the RET proto oncogene by gene rearrangement and form the thyroid tumor-specific chimeric oncogene known as PTC2.

Given the relationship of this gene with neurodegenerative diseases, this study will focus on the analysis of this gene's sequence. Additionally, the chromosome where it is localized, the 17th chromosome, will be also analyzed.

Objectives

The main objective of this study is to observe if there are significant differences in methylation chromosome 17 and the PRKAR1A between the four diseases previously introduced.

Materials & Methods

For this study, the materials used are Illumina HiSeq 2000 runs from each of the four diseases: Down's syndrome, Parkinson's disease, Dementia of Lewy bodies and Alzheimer's disease.

Due to the limited hardware capacity, only the first run of each disease will be used for the subsequent analysis; therefore the runs used are: SRR1272802 for Down's syndrome, SRR1272799 for Parkinson's disease, SRR1272795 for Dementia of Lewy bodies and SRR1272791 for Alzheimer's disease.

First the .sra files of the selected runs were obtained using the SRA Tools (version 3.1.1) command '*prefetch*' from the National Library of Medicine (NIH). After the files were downloaded, they were converted into fastq files using the command '*fastq-dump*'. This step was performed on the Windows terminal (Powershell).

The reference sequence, corresponding to chromosome 17, was downloaded from the National Library of Medicine, from the Genome assembly GRCh37.p13. This choice was driven by previous studies, where the authors compared the diseases against the reference hg19, also called GRCh37. Finally, chromosome 17 was selected and the FASTA file was downloaded (NCBI Reference Sequence: NC_000017.10). Later this reference was indexed, using bwa tools (version 0.7.17) '*bwa index'* on Ubuntu, to obtain essential files to align the different reads.

The sequences from the diseases, obtained previously in fastq format, were aligned to the hg19 chromosome 17 reference sequence using '*bwa mem*' (bwa version 0.7.17). The resulting aligned files were saved as SAM files. Next, the SAM files were converted to BAM files to allow for more efficient storage and processing. Additionally, the BAM files were

sorted by genomic coordinates and indexed to facilitate faster queries. The conversion from SAM to BAM files was done in Ubuntu using samtools (version 1.21).

An initial attempt was made to extract methylation information using Bismark (version 0.22.3). The usage of '*bismark_methylation_extractor*' caused several errors, creating empty files. These errors were likely due to incompatible file formats due to an excessive clipping of the genome sample, as the reference genome prepared with '*bismark_genome_preparation*' did not cause any problems.

A second attempt at obtaining the methylation information was made, using bedtools (version 2.28.0). A reference annotated CpG bedfile was created using the fasta reference genome. Then the methylation data was extracted using '*bedtools intersect*'; however, this command was stopped due to the RAM limitations of the hardware.

In the end, only the methylation count was obtained. Using '*samtools mpileup*', the count of methylated and unmethylated cytosines was obtained from the generated BAM files from each of the diseases.

Next, these methylation counts were exported to R software (version 2024.02.02 + 764). There, the files were processed to obtain the chromosome wide methylation levels and the methylation levels for the gene of interest (PRKAR1A, chromosome location: 68511780 - 68551319 bp). The chromosome locations from this gene were obtained from The human protein atlas website.

Results

The resulting methylation counts for the diseases were obtained using R from the methylation counts file generated using mpileup.

Methylation percentage	Down's syndrome	Alzheimer's disease	Dementia of Lewy bodies	Parkinson's disease
Chromosome 17	9.817 %	9.77 %	9.81 %	10.45 %
PRKAR1A gene	1.85 %	2.34 %	2.33 %	2.08 %

Table 1. Methylation percentages of the 4 diseases.

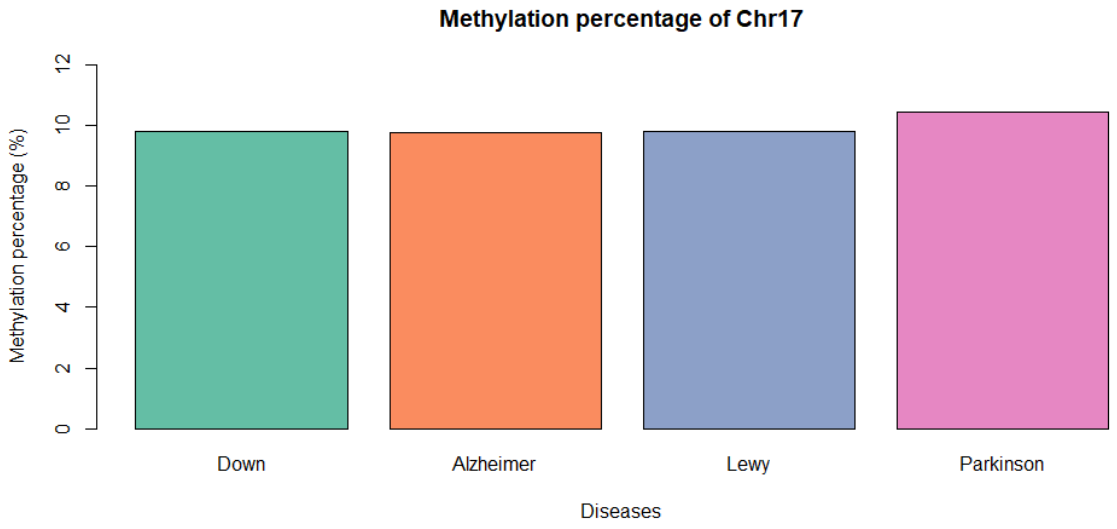

Figure 1. Graph of Methylation percentages for Chromosome 17.

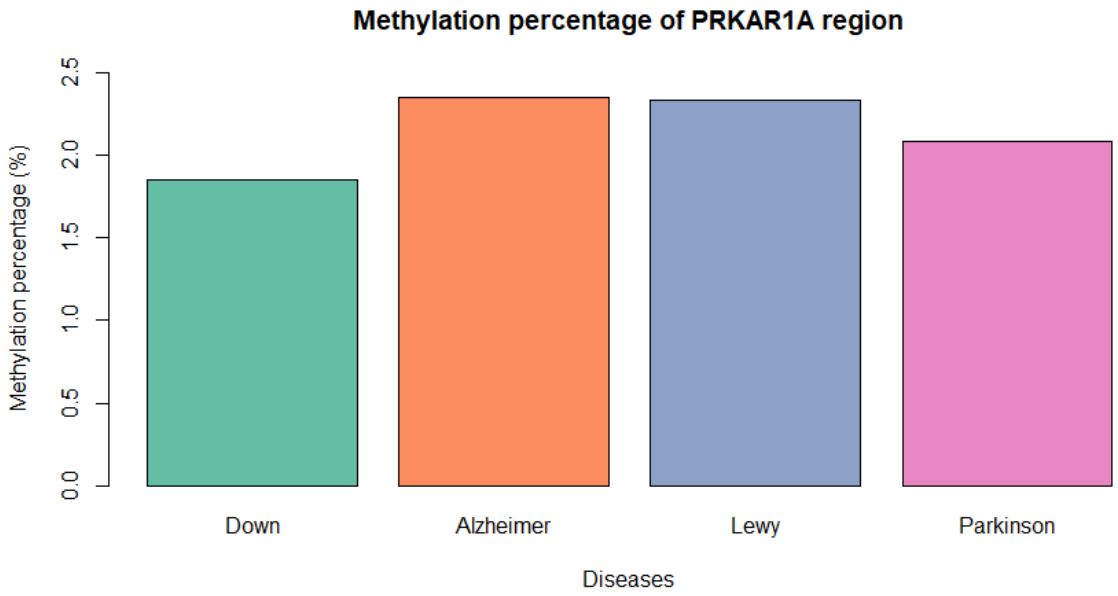

Figure 2. Graph of Methylation percentages for PRKAR1A region.

Limitations

This study has found many difficulties and limitations.

The main limitation of this study is the lack of statistically significant results. The methylation counts obtained and shown in the previous section provides only a descriptive summary of how much methylated is the region of interest. From this information, only assumptions about the relationship between the methylation levels and the diseases can be made because there is no statistical significance assessment. Additionally, since only one sample was used for each of the diseases, the biological variability in methylation levels could not be accounted for.

Another limitation was the lack of hardware strong enough. The files downloaded and obtained during the analysis and the files needed to execute the functions occupied an immense amount of space. This space surpassed the capabilities of the computer, so an external memory was acquired to advance with the project. Additionally, the *bedtools intersect* step could not be done due to limitations in the hardware RAM, which caused the code to be killed before damaging the computer.

Finally, the methylation counts of the diseases were not compared against a healthy reference. This lack of control group means that the diseases can only be compared against each other, which can lead to unreliable or biased conclusions.

Conclusion

The results of this study suggest that there is an intrinsic relationship between the protein of interest (PRKAR1A) and neurodegenerative diseases such as Alzheimer's disease, Dementia of Lewy bodies disease and Parkinson's disease. These three diseases have more methylated percentage in their PRKAR1A regions than Down's syndrome. Furthermore, Alzheimer and Dementia of Lewy bodies show a higher methylation percentage than Parkinson's disease. This similarity in a gene such as PRKAR1A, which previous studies have related to neurodegenerative diseases, is supported by their similarity in phenotype. Both diseases produce memory loss, as opposed to the main symptoms of Parkinson and Down syndrome. This similarity, both in methylation percentage and in symptomatology, indicates that this gene may regulate an important pathway that is altered in both diseases.

Additionally, Parkinson's disease shows a higher methylation percentage through the whole 17th chromosome. This higher methylation in the whole chromosome, but not in the PRKAR1A regions suggests that there may be more pathways affected by methylation in Parkinson's disease, but these pathways are different from the genes affected in the other diseases.

Finally, Down's syndrome shows a low methylation in the gene of interest, and a methylation level similar to the Dementia of Lewy bodies and Alzheimer's diseases. This may be because Down's syndrome is caused by an extra copy of chromosome 21, where we would expect to find more significant differences. However, individuals with Down syndrome suffer from developmental delay, intellectual disability, and an early-onset of neurodegeneration. This early onset of neurodegeneration and diseases similar to Alzheimer's could be the cause of the similar chromosome 17 methylation between Down's syndrome and Dementia of Lewy bodies and Alzheimer's diseases.