International Journal of Clinical & Medical Informatics

ISSN: 2582-2268 Research Article | Vol 6 Iss 1

Identification of Genomic Variants from the Transcriptome Data of Fetal Alcohol Spectrum Disorder

Shambhavi Singh, Payal Trivedi and Prachi Srivastava*

Amity Institute of Biotechnology, Amity University, Uttar Pradesh Lucknow Campus, India

Correspondence should be addressed to Prachi Srivastava, Amity Institute of Biotechnology, Amity University, Uttar Pradesh Lucknow Campus, 226028, India

Received: March 16, 2023; Accepted: April 09, 2023; Published: April 16, 2023

ABSTRACT

Fetal alcohol spectrum disorder is a group of disorders related with alcohol characterizing a wide range of loss due to the prenatal exposure of alcohol. It includes birth-related congenital anomalies including growth deficits, facial dysmorphism and malformations in multiple physiological systems. Defects caused by this disorder are irreversible. In this study, identification of genomic variants viz. SNPs, Indels etc. were performed by using RNA Seq data of Fetal alcohol spectrum disorder. Here, transcriptome data was taken from the databases viz. GEO, SRA, ENA.RNA Seq data of mice (*mus musculus*) in which six samples of FASD were selected. Quality check was performed by using Fastqc software and alignment with the reference genome was completed by using BWA mem2 software. Further, Freebayes tool was used for the detection of significant variants in the diseased samples and annotation was done by SnpEff tool. Mutational analysis was done by using the mutation taster tool. Findings reveal that two significant single nucleotide polymorphisms were detected viz. A>C, T>A, reveals the novel missense variation. These significant variations have shown polymorphic effects after mutational analysis. Hence, this gives information about the detection of significant variants focusing on SNPs. It concludes that this significant computational approach can bring new hopes and an array of light in terms of therapeutic diagnosis to cure this disorder.

KEYWORDS

Fetal alcohol spectrum disorder; Variant analysis; SNPs; Missense mutation

INTRODUCTION

Exposure to alcohol is the most common cause of environment induced birth defects. Until 1968, birth defect due to alcohol exposure was not recognized. Alcohol consumption causes prenatal defects was first clinically recognized in the year 1973 [1]. The term 'Fetal alcohol spectrum disorder' refers to a group of disorders linked to alcohol that exhibit the whole spectrum of harm caused by prenatal alcohol exposure. It includes congenital anomalies that can include growth deficits, facial dysmorphism and malformations in multiple physiological systems [2]. FASD also includes many neurodevelopmental disorders such as intellectual disabilities, microcephaly and attention deficient hyperactivity [3]. The normal physiology of the face and the brain might be disturbed by heavy alcohol use during the first trimester of pregnancy. During pregnancy,

Citation: Shambhavi Singh, Identification of Genomic Variants from the Transcriptome Data of Fetal Alcohol Spectrum Disorder. Int J Clin Med Info 6(1): 13-18.

exposure to alcohol may have an impact on healthy physiology. Three distinct facial anomalies include narrow eye apertures, a smooth space between the lip and the nose, and a thin upper lip typically appear in a pattern. Fetal alcohol syndrome defects are permanent and not reversible. Alcohol consumption is never healthy, especially for expectant mothers. Alcohol is easily transferred from a mother's bloodstream to the blood of an unborn child through the placenta. Alcohol can disrupt the normal physiology of the brain and other important organs, tissues, and physiological systems when it is present in a developing baby's bloodstream. The prevalence of fetal alcohol spectrum disorder globally is 0.77% with the highest prevalence of 2%-5% in Europe and North America [4]. It is a leading non-preventable cause of birth defects and developmental disability. The reported incidence of FASD is 0.2 - 3.0 per 1000 live births for most countries. However, cases upto 89 per 1000 live births have been reported in South Africa where production of wine is abundant. The highest rate of FASD was reported by the World Health Organization European region with 19.8 per 1000 population [5]. The diagnosis of FASD is usually challenging due to poor reliability of self-reported drinking at the time of maternity histories [6]. In this study variant analysis was performed. It is a process of analyzing and identifying the effect of particular variants on transcriptome data of FASD patients. It implies identification of variants viz. single nucleotide polymorphism (SNPs) and insertions and deletions (Indels) from next generation sequencing analysis pipelines. Here, variants were identified which cause mutational changes in the genome of an organism due to prenatal alcohol exposure. For this purpose, six mice (*mus musculus*) samples were retrieved. Raw data were preprocessed, and further variant detection was performed in which significant single nucleotide polymorphisms were identified which cause novel missense mutation in patients of Fetal alcohol spectrum disorder [7].

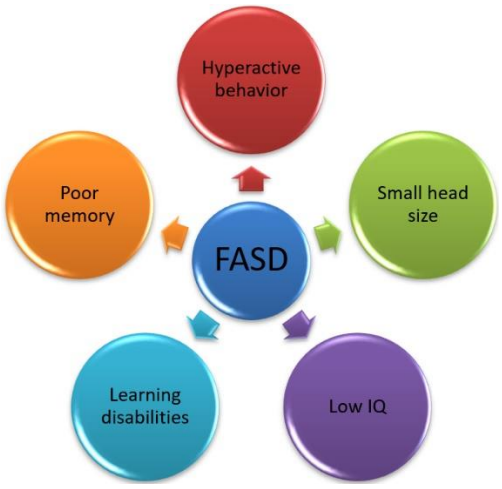

Figure 1: Diagrammatic representation of FASD symptoms.

MATERIALS AND METHODS

Data collection and Retrieval

Six samples of RNA Seq data were retrieved from the hippocampus region of brain of ethanol exposed mice from the publically available databases viz. Gene Expression Omnibus (GEO), SRA, ENA [8]. The main inclusion parameters for data selection were: (A) RNA seq data from mice (*mus musculus*) (B) Data should have at least 4-5 diseased samples (C) Data should be in FastQ format. Therefore, a dataset with GEO Id GSE184615 was selected which contains the transcriptome data of six diseased mice samples in Fastq format [9]. These Fastq files were retrieved from SRA and ENA databases for further processing of data (Table 1).

GEO Accession No.	SRA ID	Organ/tissue
GSE 184615	SRR16013004	Brain (Hippocampus)
GSE 184615	SRR16013005	Brain (Hippocampus)
GSE 184615	SRR16013006	Brain (Hippocampus)
GSE 184615	SRR160130010	Brain (Hippocampus)
GSE 184615	SRR160130011	Brain (Hippocampus)
GSE 184615	SRR160130012	Brain (Hippocampus)

Table 1: Tabular representation of transcriptome data.

Pre-Processing of Data

Quality check of the data was performed by using FastQC software. It analyses and checks the quality of raw data of

high throughput sequencing [10]. Trimmomatic software was used to trim the poor-quality sequences and retrieve the good quality sequences. Further, processing of data was proceeded after trimmomatic. Mapping of the sequences with the reference genome was done by using Bwa-mem2 software. After alignment, the output file must be in BAM format as Bwa-mem2 converts the data into BAM format in a single run [11].

Removal of Duplicates

Duplicates present in the sequences are known as redundants [12]. This redundancy may create major issues during sequencing. Therefore, they must be removed. Hence, Marking and removal of duplicates is an important step in the processing of sequenced data [13]. PCR duplicates were marked by mark duplicates tool and then removed with the help of Rmdup (version 2.0.1) tool of Picard package. It removes multiple paired reads with similar external coordinates. Rmdup only maintains the pair with highest mapping quality. Rmdup accepts aligned BAM file as input data [14].

Variants Detection

Variant detection may be identified as single nucleotide polymorphic (SNP) or Indels etc. [15]. The Free Bayes tool (version is 1.3.6) was used for variant detection. Free Bayes is a Bayesian genetic variant detector to find small polymorphisms, specifically SNPs (single-nucleotide polymorphisms) [16]. It produces VCF format dataset with interpretation of variants detected in the samples. It is used to merge multiple datasets into a single file [17].

Variants Annotation

It is a process by which variants and mutations in the genome sequences were assigned for functional information. The outcomes of such annotations are beneficial because they can directly influence the conclusions arrived in disease studies [18].

This step is necessary to analyze, identify and prioritize the variants in a study [19]. Variant annotation was completed in several steps. In first step normalization and alignment of data was done by Bcftool norms, version 1.15. It lefts align and normalize the variants. Further, VCF filter (version 1.0.0) was used to filter the VCF file in which variants were compiled by using parameters Dp value as standard parameter. Dp value gives the information about the depth of reads or no. of times the data was sequenced. Here, Dp value was taken as >2000. Finally, the annotation was done by SnpEff tool (version 4.3) which was used to annotate the functional variants and prediction of the various effects of genetic variants [20].

Mutational Analysis

Mutational analysis of the detected variants was done by mutation taster (version 2021) tool to determine whether the variation is harmful or polymorphic in nature. Mutation Taster uses a Bayes classifier [21]. The Bayes classifier computes the probability for the alteration to be either a disease mutation or a benign polymorphism based on the results of all tests and the characteristics of the alterations [22]. Here, ID of transcripts formed by the variants, position of transcript and altered base was required as input data.

RESULTS AND DISCUSSIONS

In variant analysis of the total six samples of transcriptome data of FASD, total variants analysed were 957 out of which 736 Single nucleotide polymorphisms were detected, 112 insertions, 26 deletions were detected as given below (Figure 2) but they were not shown any significant variation, so were not processed further. Out of 736 SNPs eight showed missense mutations while 14 showed silent mutations as given below in Figure 3. Silent mutations do not cause any significant variation in the genome. Hence,

missense mutation was detected as a significant variation in the patients of FASD.

Number variants by type

Type	Total
SNP	736
MNP	75
INS	112
DEL	26
MIXED	8
INV	0
DUP	0
BND	0
INTERVAL	0
Total	957

Figure 2: Total no. of variants detected.

Number of effects by functional class

Type (alphabetical order)	Count	Percent
MISSENSE	8	36.364%
SILENT	14	63.636%

Missense / Silent ratio: 0.5714

Figure 3: Represent significant variants detected.

Further, filtration was done on the basis of protein coding transcripts formed by the detected SNPs which have shown novel missense variation. Hence, two SNPs were filtered out which have formed maximum number of protein coding transcripts viz. A>C and T>A. A>C variation was detected in chromosome 11 and T>G variation was detected in chromosome 2 Amino acid glutamine was converted to alanine (Glu727Ala) and codes for a significant which was encoded by Gria1 gene whereas isoleucine amino acid was converted to asparagine (ile661Asn) which codes for crucial proteins which was encoded by Map1a gene respectively. After mutational analysis, both of the identified SNPs have shown polymorphic effect i.e. they have not shown any harmful or disease causing effect in the genome of patients. If a missense mutation exhibits polymorphism, it is regarded as probably harmless as represented in the table below (Table 2).

Chr no.	Reference	Altered	Type	Annotation	Gene	Mutation analysis
11	A	C	SNP	MISSENSE	Gria1	Polymorphism
2	T	A	SNP	MISSENSE	Map1a	Polymorphism

Table 2: Mutational analysis of the detected SNPs.

CONCLUSION

As already discussed, drinking alcohol during pregnancy always has a negative impact on the newborn and causes complex disorders like FASD. FASD generally goes untreated and unrecognized. In a clinical environment, SNP identification can be helpful for an accurate diagnosis of FASD [23,24]. Exposure to ethanol also results in persistent and irreversible mutations in which missense mutations are probably harmful to living beings. In terms of abundance in human genomes, any information related to SNPs and mutations can be understood well by transcriptome data analysis [25,26]. In this study, transcriptome data was taken to identify the variants responsible for complex disorders like FASD. Here, two significant SNPs were detected as significant variants which have shown novel missense mutation in the patients of FASD. Therefore, the method of transcriptome investigation will be very useful for the future diagnostic applications for such complex neurodevelopmental disorders [27,28]. This crucial computational approach can also provide fresh viewpoints and new information on the therapeutic identification and management of this ailment [29,30].

ACKNOWLEDGEMENTS

This is not just to follow the custom of writing acknowledgement but to express and record our heartfelt

feeling of gratefulness to all those who directly or indirectly helped us in this work. Further we wish to acknowledge high through put analysis tools for conducting this study.

DATA AVAILABILITY STATEMENT

The datasets that support the findings of this study are openly available in NCBI Gene Expression Omnibus Database at https://www.ncbi.nlm.nih.gov/geo.

COMPETING INTERESTS AND FUNDING

The authors declare that they have no conflict of interest, and no financial interest is involved in this study.

REFERENCES

1. Wagner JC, Tergeist M, Kruse B, et al. (2020) Fetal alcohol spectrum disorders in adults. The Neurologist 91: 1069-1079.
2. May PA, de Vries MM, Marais AS, et al. (2022) The prevalence of fetal alcohol spectrum disorders in rural communities in South Africa: A third regional sample of child characteristics and maternal risk factors. Alcoholism: Clinical and Experimental Research 46(10): 1819-1836.
3. Colom J, Segura-García L, Bastons-Compta A, et al. (2021) Prevalence of Fetal Alcohol Spectrum Disorders (FASD) among children adopted from eastern European countries: Russia and Ukraine. International Journal of Environmental Research and Public Health 18(4): 1388.
4. Wozniak JR, Riley EP, Charness ME (2019) Clinical presentation, diagnosis, and management of fetal alcohol spectrum disorder. The Lancet Neurology 18(8): 760-770.
5. Maya-Enero S, Ramis-Fernández SM, Astals-Vizcaino M, et al. (2021) Neurocognitive and behavioral profile of fetal alcohol spectrum disorder. Annals of Pediatrics 95(3): 208.e1-208.e9.
6. Fainsod A, Bendelac-Kapon L, Shabtai Y (2020) Fetal alcohol spectrum disorder: Embryogenesis under reduced retinoic acid signaling conditions. The Biochemistry of Retinoid Signaling III: Vitamin A and Retinoic Acid in Embryonic Development 95: 197-225.
7. Nayak RB, Murthy P (2008) Fetal alcohol spectrum disorder. Indian pediatrics 45(12): 977-983.
8. de Sena Brandine G, Smith AD (2019) Falco: High-speed FastQC emulation for quality control of sequencing data. F1000Research 8: 1874.
9. Brown J, Pirrung M, McCue LA (2017) FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics 33(19): 3137-3139.
10. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114-2120.
11. Jung Y, Han D (2022) BWA-MEME: BWA-MEM emulated with a machine learning approach. Bioinformatics 38(9): 2404-2413.
12. Pireddu L, Leo S, Zanetti G (2011) SEAL: A distributed short read mapping and duplicate removal tool. Bioinformatics 27(15): 2159-2160.
13. Ebbert MT, Wadsworth ME, Staley LA, et al. (2016) Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. BMC Bioinformatics 17: 491-500.

14. Fu Y, Wu PH, Beane T, et al. (2018) Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. BMC Genomics 19: 1-14.
15. Liang Y, Lin H, Zou L, et al. (2022) Rapid detection and tracking of Omicron variant of SARS-CoV-2 using CRISPR-Cas12a-based assay. Biosensors and Bioelectronics 205: 114098.
16. Clevenger J, Chavarro C, Pearl SA, et al. (2015) Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations. Molecular Plant 8(6): 831-846.
17. Liu J, Shen Q, Bao H (2022) Comparison of seven SNP calling pipelines for the next-generation sequencing data of chickens. Plos One 17(1): e0262574.
18. Ulintz PJ, Wu W, Gates CM (2019) Bioinformatics analysis of whole exome sequencing data. Chronic Lymphocytic Leukemia: Methods and Protocols 1881: 277-318.
19. Cingolani P (2012) Variant annotation and functional prediction: SnpEff. In Variant Calling: Methods and Protocols 2493: 289-314.
20. Cingolani P, Platts A, Wang LL, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w^{1118}; iso-2; iso-3. Fly 6(2): 80-92.
21. Kong X, Zhong X, Liu L, et al. (2019) Genetic analysis of 1051 Chinese families with Duchenne/Becker muscular dystrophy. BMC Medical Genetics 20(1): 1-6.
22. Ni SH, Zhang JM, Zhao J (2022) A novel missense mutation of CRYBA1 in a northern Chinese family with inherited coronary cataract with blue punctate opacities. European Journal of Ophthalmology 32(1): 193-199.
23. Autti-Rämö I, Fagerlund Å, Ervalahti N, et al. (2006) Fetal alcohol spectrum disorders in Finland: Clinical delineation of 77 older children and adolescents. American Journal of Medical Genetics Part A 140(2): 137-143.
24. Doak J, Katsikitis M, Webster H, et al. (2019) A fetal alcohol spectrum disorder diagnostic service and beyond: Outcomes for families. Research in Developmental Disabilities 93: 103428.
25. Flannigan K, Wrath A, Ritter C, et al. (2021) Balancing the story of fetal alcohol spectrum disorder: A narrative review of the literature on strengths. Alcoholism: Clinical and Experimental Research 45(12): 2448-2464.
26. Popova S, Charness ME, Burd L, et al. (2023) Fetal alcohol spectrum disorders. Nature Reviews Disease Primers 9(1): 11.
27. McCarthy R, Mukherjee RA, Fleming KM, et al. (2021) Prevalence of fetal alcohol spectrum disorder in Greater Manchester, UK: An active case ascertainment study. Alcoholism: Clinical and Experimental Research 45(11): 2271-2281.
28. Okurame JC, Cannon L, Carter E, et al. (2022) Fetal alcohol spectrum disorder resources for health professionals: A scoping review protocol. BMJ Open 12(9): e065327.
29. May PA, de Vries MM, Marais AS, et al. (2022) The prevalence of fetal alcohol spectrum disorders in rural communities in South Africa: A third regional sample of child characteristics and maternal risk factors. Alcoholism: Clinical and Experimental Research 46(10): 1819-1836.
30. Ceccanti M, Fiorentino D, Coriale G, et al. (2014) Maternal risk factors for fetal alcohol spectrum disorders in a province in Italy. Drug and Alcohol Dependence 145: 201-208.