Original article

In Silico Identification of potential Drug Target and Analysis of Effective Single Nucleotide Polymorphisms for Autism Spectrum Disorder

Manisha Ray*, Jagruti Mishra, Abhishiktta Priyadarshini, Souvagyalaxmi Sahoo

Tectona Biotech Resource Center, Bhubaneswar-751002, Odisha.

1. Introduction

Autism Spectrum disease is a really heritable neuro developmental sickness that impacts the brain, the immune device, the gastrointestinal tract and other organ structures [1, 2]. It characterized with the aid of restricted interests, repetitive behaviour, impairments in social communication etc [3, 4]. Autism is widely taken into consideration to be a multi factorial sickness that effects from genetic in addition to non genetic hazard elements, but the exact cause of autism is unknown, however a strong genetic component has been identified from a family and twin studies. Moreover genetic studies discovered that alternation inside the developmental pathways of neuronal and axonal systems are strongly involved in synaptogenesis emerge from single gene mutations [5]. ASD is a highly prevalent neuro disease condition with no current treatment available. As experimental proof emerges in current years, advances in genetics and genomics have identified autism risk genes [6]. Complex genetic interactions appear responsible for a high degree of heterogeneity of the clinical symptoms in ASD. The identification of the genetic components of autism spectrum disorder has advanced rapidly in current years, particularly with the demonstration of de novo mutations as a critical source of causality [7].

Progress in genomics and the associated technological, statistical and bioinformatics advances have facilitated the successful implementation of genome-wide association studies (GWAS) towards understanding the genetic foundation of common diseases [8]. Genome wide association research on individuals with ASD and their families revealed several threat genes that may be the common molecular targets in Autism. However, the list of autism threat genes affords us a path to understand the potentially more prone pathways in neurons that may be therapeutic objectives to develop extra efficient interventions for ASD [9]. There were many genes recognized as target genes and reported to cause ASD with mutation including, GABRG3 [10], MTOR [11], FMRP [4], CHD8, EHMT1, SATB2, DLX1 [12], FMR1, TSC1/2, MECP2, CHD2, ARID1, TBR1, DYRK1A, GRIN2B [13], GABRB3, CC2D1A, UBE3B, AMT, PEX7, VPS13B [14], PTEN, CNTNAP2, SLC9A9, PEX7, CC2D1A, BCKD K, SYNGAP1, DYRK1A, SCN2A [6] etc. The considerable advances in genetic and genomics study has been rapid, more work is needed to fully understand the capability for impacting disease treatment and prevention, specially the mental health [5]. Thus to satisfy the unconventional treatment of genetic diseases, the Genome wide association research (GWAS) with hundreds of lots of SNPs are popular techniques which reveal the genetic basis of human complex disorders [16].
Single nucleotide polymorphisms (SNPs) are the most common form of source of variations in the human genome. Among all of the SNPs, nonsynonymous SNPs (nsSNPs) are vital as those can modify the amino acid residues and may create damage or modify protein coding sites and those effects or deleterious results of SNPs are generally attributed to their impact on the protein structures and functions [17, 18].

Thus, this present study retrieved some vital target genes for Autism Spectrum Disorder and analyzed the associated deleterious nsSNPs in account of having a new path to deal with and cure the Autism Spectrum Disorder by following the related tools and databases of Genome Wide Association Study, such as human disease gene databases like Autdb, GeneCards, MalaCards, additionally the use of STRING database became worried to conclude the target gene identification manner. Then the SNP retrieval was completed by way of using SIFT, PolyPhen2, PROVEAN and SNAP2, followed by identified target genes to carry out the whole study by figuring out the most deleterious nsSNPs, which results the novel drug target with most effective SNPs for the development of latest drugs towards ASD.

**Material and Methods**

**AutDB**

AutDB (http://autism.mindspec.org/autdb/Welcome.do) is an evolving modular database for the autism research network that is targeted on genes implicated in autism susceptibility. Thus this examines retrieved related Autism genes and achieved for further evaluation.

**GeneCards**

GeneCards is also known as Human Gene Database (https://www.genecards.org/), is routinely integrates gene-centric information from ~a hundred twenty five net sources, which encompass genomic, transcriptomic, proteomic, genetic, clinical and practical information. Here, the Autism associated genes were retrieved from GeneCards database via key-word search.

**MalaCards**

MalaCards (https://www.malacards.org/) is an included database of human sickness and problems. It also analysing disease associated gene units through Gene Analytics to yield affiliated pathways, phenotypes, compounds, and GO terms. Here on this database the amassed target genes of autism have been confirmed with MalaCard ID and the disease class.

**STRING**

STRING (https://string-db.org/) is the practical protein association community, also a database of recognized and anticipated protein-protein interactions. So, because of its potentiality of evaluation, the collected Autism genes had been interpreted through functionally community via STRING.

**dbSNP**

The dbSNP (https://www.ncbi.nlm.nih.gov/snp/?term) database, established through National Centre for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) serve as a crucial repository for each single base nucleotide substitutions and quick deletion and insertion polymorphisms, which emphasizes the gathering of SNPs for amassed target genes of Autism to complete this present work.

**SIFT**

After the collection of target SNPs those were subjected to predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids through SIFT (http://sift.jcvi.org/).

**PolyPhen-2**

The tested SNPs through SIFT these were once more tested via PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) as it predicts feasible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations.

**PROVEAN**

PROVEAN (http://provean.jcvi.org/index.php) is likewise called Protein Variation Effect Analyzer, that is useful for filtering sequence versions to become aware of non-synonymous or indel variations which can be expected to be functionally important, so the established SNPs after PolyPhen-2 evaluation have been analyzed in PROVEAN and consequences have been achieved for further observance.

**SNAP2**

The demonstrated SNPs from PROVEAN have been once more analyzed in SNAP2 (https://www.rostlab.org/services/SNAP/) device as it has the set of rules to are expecting the impact or effect of single amino acid substitutions on protein function.

**Results**

**Retrieval of Target Genes**

There are a number of genes are involved in different ways to cause the autism spectrum disorder or Autism. So this presented work was gone through to search the autism associated target genes from distinctive databases like Autdb, GeneCards and MalaCards and so forth. The AutDB is a molecular database for autism research, which contained 6321 numbers of genes of Autism, however in those genes lists there had been many repetitions of genes had been befell, so by eliminating the repeat genes there had been total 856 numbers of target genes for autism were accumulated from AutDB. Thereafter the collection of target genes from AutDB, another human disease gene database GeneCards, was taken in to attention and looked for the autism target genes, which gave 4390 numbers of genes for autism.
Then there was a comparison made between the collected target genes of autism from these above two considered databases, i.e. the common genes from both databases were screened out and results 482 numbers of genes, which were taken for another analysis like validation process in MalaCards database. As MalaCards is an integrative database of human diseases, the 482 amassed genes for autism from AutDB and GeneCards had been tested via MalaCards i.e. from 482 genes, which genes have been also present as target to cause autism in MalaCards were once more screened out and eventually were given 195 numbers of target genes for autism (Fig. 1(A)).

Functional Network Analysis

The collected 195 numbers of target genes were subjected to network analysis through STRING, which enrich the functional network between the 195 genes based on highest confidence score for interaction 0.900 and resulted 86 numbers of genes (Fig. 1(A), Fig. 2), were distinctly interacted functionally and subjected for further analysis. The built network contains 189 numbers of nodes, 182 numbers of edges and the PPI enrichment value was turned into < 1.0e-16.

Collection of SNP data

The data set of SNPs for 86 numbers of target genes of Autism have been gathered from database of SNP or dbSNP based totally on some criteria like non synonymous SNPs with having protein annotation, pathogenic and/or benign clinical significance, functional annotations including 3'UTR, 5'UTR, frameshift, stop gained, missence, nonsense etc and validation status by using 1000 genome, which provides 31 numbers of genes with total 94 numbers of nsSNPs for Autism among 86 target genes (Fig. 1(B)).

Functional Mutation Analysis of Target Genes

For the accuracy of computational techniques, prediction of maximum deleterious SNPs can be extended by means of the usage of a mixture of numerous exceptional in silico algorithms through different tools. Therefore by means of furnishing the accuracy to predict the most deleterious or damaged SNP right here a few tools had been used like SIFT, which predicts only 167 nsSNPs of 31 target genes from 94 nsSNPs of 31 genes respectively as tolerated and deleterious followed by the prediction scores like; prediction score with > 0.05 was taken to be tolerated and score with < 0.05 was considered to be deleterious, according to their functional influence of amino acid substitution; but in the present study only the deleterious nsSNPs were taken into consideration for further analysis, so only SIFT predicted 17 deleterious nsSNPs of 11 (SHANK3, DLG4, NRXN1, CACNA1D, LRP2, TSC2, DRD2, AMPD1, ADSL, LAMB1, GPHN) (Table 1) drug target genes were subjected to validation through PolyPhen 2, which result 9 nsSNPs for 6 (CACNA1D, LRP2, TSC2, DRD2, AMPD1, ADSL) (Table 1) genes with possible damaging out of 17 nsSNPs by means of the use of some algorithms and prediction scores i.e. 0.000 indicates probably benign and 0.999 indicates probably damaging. Further those 9 nsSNPs had been once more subjected to analyze in PROVEAN tool to screened out the high risk SNP as either with neutral or deleterious effects, and this server predicted 4 deleterious effect nsSNPs for 2 genes (CACNA1D, AMPD1) (Table 1) out of those 9 on the basis of its screening criteria i.e. the prediction value with > -2.5 is shows neutral effect and the value with < -2.5 shows deleterious effects on protein. Then at last those 4 nsSNPs submitted to SNAP2 tool to examine the impact of amino acid variation which gave one final high risk 3 nsSNP with rsIDs rsID35859650, rs34526199 and rs61752478 as effective on target gene AMPD1 to cause Autism Spectrum Disorder (Fig. 1(B)), (Table 1).

Table 1: Predicted nsSNPs with respective prediction values and effects from SIFT, PolyPhen2, PROVEAN and SNAP2 for collected Target Genes of ASD

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>Gene Names</th>
<th>rsID</th>
<th>SIFT</th>
<th>PolyPhen2</th>
<th>PROVEAN</th>
<th>SNAP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SHANK3</td>
<td>rs16729471</td>
<td>0.033</td>
<td>Benign</td>
<td>Benign</td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
<td>Benign</td>
<td>Benign</td>
<td>1.434</td>
</tr>
<tr>
<td>2</td>
<td>DLG4</td>
<td>rs269934586</td>
<td>0.004</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>4</td>
<td>CACNA1D</td>
<td>rs41276445</td>
<td>0.027</td>
<td>Probably damaging</td>
<td>Benign</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.074</td>
<td>Benign</td>
<td>1.000</td>
<td>2.795</td>
</tr>
<tr>
<td>5</td>
<td>LRP2</td>
<td>rs2281589</td>
<td>0.033</td>
<td>Possibly damaging</td>
<td>Possibly damaging</td>
<td>0.897</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.924</td>
<td>Possibly damaging</td>
<td>0.857</td>
<td>NA</td>
</tr>
<tr>
<td>6</td>
<td>TSC2</td>
<td>rs45537352</td>
<td>0.027</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.029</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>7</td>
<td>DRD2</td>
<td>rs1B01020</td>
<td>0.035</td>
<td>Probably damaging</td>
<td>0.992</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.993</td>
<td>Probably damaging</td>
<td>0.993</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.597</td>
<td>Neutral</td>
<td>0.38</td>
<td>(66%)</td>
</tr>
</tbody>
</table>
In the field of computational genomics research, determination of high-risk nsSNPs which may induce disease associated phenomena is getting importance. So this present study involved identification of important drug targets of autism from different human disease databases and uses multiple computational methods to detect the high-risk pathogenic mutation in the important target genes of autism by analysing single nucleotide polymorphism.

Thus as the resulted data mentioned above that total 482 numbers of common target genes for autism were screened out by comparing the two databases i.e. Autdb, that encompasses detailed annotation of both rare and common genetic variants associated with ASD for candidate gene prioritization [19] and GeneCards, a web based compendium of human disease associated genes. These genes were validated again through MalaCards database which is effectively addresses some of the major challenges facing disease bioinformatics i.e. the gene-disease relationship [20] and got 195 final target genes. The overall retrieval process for target genes was done to get more reliable, important and more authenticated targets which might be helpful for further drug design research. Further after the database search the genes were again validated through functional network association in STRING, which is one of the repositories and provides opportunity for building physical and functional networks between related genes and/or proteins as per the gene ontology such as molecular function, biological process and cellular component [21], thus because of its potentiality towards the generation and accuracy of functional networks between the genes or proteins this study use the functional network generated through STRING and got 86 genes out of 195.

Fig. 2

In the field of computational genomics research, determination of high-risk nsSNPs which may induce disease associated phenomena is getting importance. So this present study involved identification of important drug targets of autism from different human disease databases and uses multiple computational methods to detect the high-risk pathogenic mutation in the important target genes of autism by analysing single nucleotide polymorphism.

Thus as the resulted data mentioned above that total 482 numbers of common target genes for autism were screened out by comparing the two databases i.e. Autdb, that encompasses detailed annotation of both rare and common genetic variants associated with ASD for candidate gene prioritization [19] and GeneCards, a web based compendium of human disease associated genes. These genes were validated again through MalaCards database which is effectively addresses some of the major challenges facing disease bioinformatics i.e. the gene-disease relationship [20] and got 195 final target genes. The overall retrieval process for target genes was done to get more reliable, important and more authenticated targets which might be helpful for further drug design research. Further after the database search the genes were again validated through functional network association in STRING, which is one of the repositories and provides opportunity for building physical and functional networks between related genes and/or proteins as per the gene ontology such as molecular function, biological process and cellular component [21], thus because of its potentiality towards the generation and accuracy of functional networks between the genes or proteins this study use the functional network generated through STRING and got 86 genes out of 195.

Fig. 2

In the field of computational genomics research, determination of high-risk nsSNPs which may induce disease associated phenomena is getting importance. So this present study involved identification of important drug targets of autism from different human disease databases and uses multiple computational methods to detect the high-risk pathogenic mutation in the important target genes of autism by analysing single nucleotide polymorphism.

Thus as the resulted data mentioned above that total 482 numbers of common target genes for autism were screened out by comparing the two databases i.e. Autdb, that encompasses detailed annotation of both rare and common genetic variants associated with ASD for candidate gene prioritization [19] and GeneCards, a web based compendium of human disease associated genes. These genes were validated again through MalaCards database which is effectively addresses some of the major challenges facing disease bioinformatics i.e. the gene-disease relationship [20] and got 195 final target genes. The overall retrieval process for target genes was done to get more reliable, important and more authenticated targets which might be helpful for further drug design research. Further after the database search the genes were again validated through functional network association in STRING, which is one of the repositories and provides opportunity for building physical and functional networks between related genes and/or proteins as per the gene ontology such as molecular function, biological process and cellular component [21], thus because of its potentiality towards the generation and accuracy of functional networks between the genes or proteins this study use the functional network generated through STRING and got 86 genes out of 195.
After the completion of target identification process this study gone through the collection of single nucleotide polymorphisms (SNPs) of related target genes to isolate the most deleterious or highly risk SNP, because the common genetic mutation is the single nucleotide polymorphism in human and about 93% SNPs are present in human genes and nSNPs have more effect on proteins [22], so this search used dbSNP of NCBI [23] to collect the nSNPs for selected target genes of Autism, as dbSNP includes disease-causing clinical mutations as well as neutral polymorphisms, [24]. But out of 86 numbers of target genes only 31 numbers of target genes were having 94 numbers of SNPs, which were non synonymous in nature were observed from dbSNP and considered for next process.

On the completion of nSNPs retrieval, those subjected to validation process on account to analyze their effects or impact on amino acid substitution or on protein structural modifications. Thus the validation process comes to find out the most deleterious or damaged nSNP which may have impact on the structural integrity of human ASD target proteins, because the deleterious effects of SNPs are generally attributed to their impact on the protein structure and function, [17] which leads to cause disease. So here, this process initiated with SIFT as it is the most well recognized tool for prediction of disease causing nSNPs due to its high performance and its easy applicability to large datasets, and is specialized for analysing disruption by nSNPs of conserved sequences [25], hence the analysis through SIFT result 17 numbers of deleterious as well as tolerated nSNPs for 11 numbers of target genes by following the scoring system. Thereafter the SNP analysis turned into validated through the Polyphen2 mediated analysis as it can predicts the phenotypic changes in proteins caused by nSNP and predicted 9 deleterious nSNPs out of 17 and 6 target genes out of 11, then out of 9 nSNPs only 4 were predicted as deleterious for 2 numbers of genes again in Provean and lastly out of these 4 only 3 predicted as highly deleterious from SNAP2 analysis and considered to be the most effective and deleterious nSNPs for the target protein AMPD1 because it was examined that the resulted nSNPs of AMPD1 were commonly predicted as deleterious and effective in all analyses, so AMPD1 might be a potential drug target to cure Autism Spectrum Disorder in future aspects.

Conclusion

Because of the shortage of medications and remedies for autism spectrum disease, this study might be the path finder for novel ideas to cure ASD, by way of identifying considerable nSNP in target genes via computational techniques. This present analysis screened total 31 numbers of potential target genes for Autism with 94 numbers of related nSNPs, which have been more demonstrated through diverse computational methods and in the end concluded with one putative drug target AMPD1 with significant nSNPs i.e. rsID335859650, rs34526199 and rs61752478. These findings obtained from the analysis could pave the manner for presenting beneficial information to the researchers and can play a vital role in computational drug layout in addition to in pharmaceutical studies area for future works.

Acknowledgement

We really thanks to Tectona Biotech Resource Centre (TBRC), Odisha, for providing necessary facilities, as well as we present our heartfelt gratitude to Dr. Shovan Kumar Mishra, Director of TBRC, for his endless support throughout this work.

6. References


Copyright 2019 BioMedSciDirect Publications IJBMR - ISSN: 0976-6685. All rights reserved.