Title: Bioinformatics Investigating Suggests Contribution of Other Chromosomes Beside Chromosome 21 In Down Syndrome Risk

Running Title: Interaction Network Analysis Revealed Key Genes in Down Syndrome

Authors: Mona Zamanian Azodi¹, Mostafa Rezaei Tavirani²*, Majid Rezaei Tavirani³, Mohammad Rostami-Nejad⁴

1- Student Research Committee, Proteomics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2- Proteomics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

3- Faculty of Medicine, Iran University of Medical Sciences, Tehran, Iran

4- Gastroenterology and Liver Diseases Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

*Corresponding Author: Mostafa Rezaei Tavirani (Tavirany@yahoo.com) Tel: 02122439787

To appear in: Basic and Clinical Neuroscience

Received date: 2018/12/15

Revised date: 2019/04/27

Accepted date: 2019/01/14

This is a “Just Accepted” manuscript, which has been examined by the peer-review process and has been accepted for publication. A “Just Accepted” manuscript is published online shortly after its acceptance, which is prior to technical editing and formatting and author proofing.
Basic and Clinical Neuroscience provides “Just Accepted” as an optional and free service which allows authors to make their results available to the research community as soon as possible after acceptance. After a manuscript has been technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as a published article. Please note that technical editing may introduce minor changes to the manuscript text and/or graphics which may affect the content, and all legal disclaimers that apply to the journal pertain.

Please cite this article as:

DOI: http://dx.doi.org/ 10.32598/bcn.9.10.410
Highlights

1. RHOA, FGF2, FYN, and CD44 are the significantly up-regulated hub-bottlenecks in down syndrome patients.
2. Biological process related to these top four genes could encountered disruptions.
3. The introduced central genes could be considered as potential biomarkers after conducting validation studies.
4. Other chromosomes rather than chromosome 21 contribute to down syndrome risk.

Plain Language Summary

Down syndrome as a genetic disorder is in a great attention for molecular studies. One way is to study down syndrome via bioinformatics. In this study, gene expression profile from microarray study was selected for more investigations. The study of down syndrome patients shows that there are certain genes with differential expression and network centrality properties. These genes are introduced as RHOA, FGF2, FYN, and CD44 that their level of expression is high in the patients. This study suggests that other than chromosomes 21, there are additional contributing chromosomes in down syndrome risk. In addition, these genes could be used for clinical studies after more analysis.
Abstract

Down syndrome (DS) is a common genetic disorder that molecular investigation could provide a better understanding of this complicated condition. In this regard, the present bioinformatic study is aimed to reveal important genes and their associated biological processes in DS via transcriptome microarray study (GSE5390) in Gene Expression Omnibus database (GEO). The differentially expressed genes (DEGs) between adult DS patients and healthy individuals from brain tissue were acquired and screened via GEO2R online software. Then, Cytoscape and its plug-ins constructed and analyzed a protein-protein interaction (PPI) network of significant DEGs. The findings indicate that there are four key possible biomarkers in DS in terms of aberrant expression and centrality in protein-protein interaction (PPI) network. Additionally, the high expression of these candidate biomarkers, including RHOA, FGF2, FYN, and CD44 could stimulate disruption in their related biological processes. Overall, although far from complete, our mapping provided a new insight to interactome level in adult DS patients by introducing four potential targets and their enriched biological features. These findings indicate that the critical aspects of DS could be affected by the products of various chromosomes (Chrs) besides Chr 21.

Keywords: Down syndrome, Protein Interaction Maps, Transcriptome, Differentially Expressed Genes, Biological Process
Introduction

Trisomy 21 is the most frequent chromosomal aberration in up to 1 in 700 population of newborns (Busch et al., 2005). Vast abnormalities in patients with this syndrome exist for instance, metabolism dysfunction, dysmorphic characteristics, and mental retardation as the most highlighted one (Bajo, Fruehauf, Kim, Fountoulakis, & Lubec, 2002; Starbuck, Cole III, Reeves, & Richtsmeier, 2017). There are different detection approaches for prenatal stage of this developmental disorder each with its pros and cons (Busch et al., 2005). Most of the molecular studies are also in this regard either via amniotic fluid or serum sample of pregnancies with trisomy 21 for diagnosis purposes (Busch et al., 2005; Tsangaris et al., 2006). However, to understand what triggers the broad range of abnormal phenotype in DS (Teeling et al.), it is required to explore the molecular profile of patients with this defect. The mechanisms behind the phenotype of DS remained to be studied via different molecular approaches (Lockstone et al., 2007). Over the past decade, there are some investigations that shed a light on complex mechanism of DS through large-scale analysis including transcriptomic, and proteomic studies (Bajo et al., 2002; Di Domenico et al., 2014; Lockstone et al., 2007). At these scales, information related to the expression profile of genes and proteins could be retrieved for any condition such as abnormal ones like a disease state (Rezaei–Tavirani, Tavirani, & Rostami, 2018). In addition, bioinformatics can reveal different aspects of identified molecules by high throughput data analysis (Rezaei–Tavirani, Bashash, et al., 2018). The function of biomolecules are handled and mediated by interacting with other molecules (Zamanian Azodi et al., 2018). One of the identified novel processes related to DS mechanism was the oxidative stress process, which was repeatedly pinpointed by many proteomic studies (Butterfield, Di Domenico, Tramutola, Head, & Perluigi, 2017; Di Domenico et al., 2014; Perluigi et al., 2011). Furthermore, this process associates with the transition of DS to DS with Alzheimer disease (Butterfield et al., 2017). Other molecular studies such a bioinformatics approach expressed that down syndrome critical region (DSCR) of chromosome 21 may have a great regulatory impact in this disorder (Chen et al., 2018). Another network analysis that was also conducted on the current and in the
combination of other microarray database indicated that there some DEGs including BCL2, HSP90 beta, UBX2, and TMEM50B that might be important in DS (Zhao, Zhang, Ren, Zong, & Kong, 2016). What is more, there are important points such as centrality properties in an interacting profile (as a scale free network) that influence the whole system. Centrality analysis is through identification of two common parameters including degree and betweenness centrality. Any changes in central elements of a scale free network could result in an abnormal phenotype such as a disease state (Zamanian-Azodi, Rezaei-Tavirani, Rostami-Nejad, & Tajik-Rostami, 2018). Therefore, identification of these essentials could provide further information and validation of biomarkers linked to the disorder condition. In this respect, protein-protein interaction network analysis was applied in this study to possibly reveal this aspect of molecular features in a disorder such as DS.

Materials and Methods
In this work, we are investigating differential expressed genes (DEGs) with centrality features in a protein-protein interaction network pattern. For this aim, at first we queried the GEO Database at the National Centre for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/geo/) for a study that provides a gene expression profile of DS patients in comparison with healthy controls. A research carried out by Bahns, et.al. entitled” Expression profiling of human adult postmortem brain tissue from adult Down syndrome and healthy control subjects”, Series GSE5390, platform: GPL96 (Lockstone et al., 2007) was selected for our bioinformatic evaluation. The chip used in this study was Affymetrix HG-U133A GeneChips. The conductors of this study used RNA extraction from postmortem brain samples of eight healthy subjects versus seven Down Syndrome types (dorsolateral prefrontal cortex) from female and male via microarray. The subjects mean ages for disease and control groups were 58.6 ± 9.4 years and 47.8 ± 10.8 years, respectively. At first, GEO2R online analyzer (https://www.ncbi.nlm.nih.gov/geo/geo2r/) in GEO Database analyzed these two groups. However, prior to that, cross comparison via box plot analysis was carried out to determine the quality of groups. If the samples are median-centered then the procedure could be continued by identification of top 250 ranked DEGs in DS. The fold change threshold was obtained by evaluating the cumulative frequency of mRNA-expressed profile between groups of
the study among 250 genes. A cut off of 2 for fold change (FC) was designated to detect the most DEGs among 250 tops. The correction test for the raw p-Value was set to Benjamini & Hochberg (False discovery rate) as the default option. The significance threshold that is acceptable for the DEGs is adj.p.value ≤ 0.05. The uncharacterized DEGs among the analyzed significant DEGs were excluded and the remained ones were applied for construction of a PPI network via Cytoscape v. 3.6.1 (https://cytoscape.org/) and its plug-in Search Tool Retrieval of Interacting Genes (STRING DB) (https://string-db.org/) (Shannon et al., 2003; Szklarczyk et al., 2016). Network Analyzer was used to determine the most critical nodes in the network in terms of interaction pattern (Assenov, Ramírez, Schelhorn, Lengauer, & Albrecht, 2007). Two parameters were fundamentals for this purpose including degree (K) and betweenness centrality (BC). Nodes with highest rank of these two features are called hub-bottlenecks (Safari-Alighiarloo, Rezaei-Tavirani, Taghizadeh, Tabatabaei, & Namaki, 2016). Following determination of hub-bottlenecks, CluePedia (http://www.ici.upmc.fr/cluepedia/) explores gene/protein/miRNA associations linked to the ClueGO networks with a designated scores. In here, it examined the expression profile of these central nodes via extracting from the expression input file in CluePedia panel. Depends on how many spots are available in the related GEO input file from for the corresponding gene, all will be extracted and merge as node labels (Bindea, Galon, & Mlecnik, 2013). Consequently, the most significant differentially expressed ones were then searched against ClueGO for enrichment and action type analysis (Bindea et al., 2013; Bindea et al., 2009). ClueGO is an application in Cytoscape that could analysis the functional properties of the queried genes. The enrichment analysis includes biological process (BP), cellular component (CC) and molecular function (MF) evaluations. In which here the BP is examined for the significant DEGs. The action type analysis between these genes were conducted via STRING Action File, V10.5., 2017., for five interaction enrichments including expression, activation, inhibition, post-translational modifications, and catalysis. To determine these interactions, kappa scoring (0-1) was used for scaling the edge strength, which is displayed by edge thickness. Moreover, this plug-in can be used for gene ontology (biological process, molecular function, and cellular component) analysis. In a way that, a network view of groups of terms associated with individual gene ontology could be provided. Our experiment covers one of the domains introducing biological process relationships. The term grouping for biological process strength was calculated by kappa statistic (this score is between 0-1) and the assigned
cutoff was 0.5. This statistic measurement is used for determination of grouping terms. The higher the k score, the lower the chance of grouping biological terms. Moreover, grouping terms are based on the default setting as follow: number of genes in terms: 3 and percentage of genes contributing in terms: 4. Likewise, default options was applied for the levels of ontology as follow: Min level of ontology: 3, max level of ontology: 8. The correction method in this regard is Bonferroni step down ($p \leq 0.05$). The enrichment/depletion test: two-sided (enrichment/depletion) based on hypergeometric

**Results**

Transcriptome analysis of DS versus healthy controls was from GEO Database via GEO2R. This comparison at first was analyzed by boxplot to evaluate the quality of samples in gene expression (see figure 1).

![GSE5390/GPL96, selected samples](image_url)
The top 250 DEGs were determined and then those with gene names and fold change ≥ 2 were considered for network establishment by STRING. The range of significance level was from 0.0003 to 0.0026. A number of 65 up-regulated and 3 down-regulated genes (FC>2) were obtained. A network of up-regulated and down-regulated genes with the addition of 100 surrounding interacting ones was constructed. The network included a main connected component (contain 156 nodes and 3420 edges) and 12 ones that were not connected to the main network (the data is not shown). For hub-bottleneck identification, the 20% of top genes with highest amount of degree and betweenness centrality were recognized. A number of 19 common genes from 34 top ones based on K and BC, were assigned as hub-bottlenecks (see Table S1). To visualize and explore the expression profile of the hub-bottlenecks DS network, CluePedia was applied (see figure 2). As depicted in the figure 2, except MBOAT4 (the low expressed gene) the other genes are differentially expressed in DS patients; however, only RHOA, FGF2, FYN, and CD44 hub-bottlenecks are significant in differential expression. (The criteria for assigning genes with more than one spots as significant differential expression ones was to have at least one spot with significant differential expression).

Statistical and centrality properties of these four significant DEGs are tabulated in the table 1. In addition, the same plug-in, CluePedia was used to gain further knowledge about the four significant DEGs in terms of interaction type (see figure 3). Activation, inhibition, Catalysis, expression, and Binding actions were determined.

Biological process analysis related to the RHOA, FGF2, FYN, and CD44 DEGs was done by CluePedia and CluGO. As it is shown in the figure 4, 15 terms are identified which are clustered in the 4 classes.
Figure 2. Normalized expression showing as node labels for 18 found genes using DS profile.

In the figure 2, the data is extracted from input file, GSE5390_series_matrix merging in CluePedia application. The rows are the number of genes’ spots in the sample that are considered as differentially expressed (DE) probes. The color scheme from red (Maximum positive Expression) to green (Maximum negative expression) shows the pattern of positive to negative expression changes. White and grey refer to zero expression and missing values, respectively.
Several spots were available for FGF2, FRN, CD44 whereas one spot was extracted for RHOA.

**Table1.** The list of significant hub-bottlenecks that are differentially expressed in DS patients. Chr refers to chromosomal location of the genes.

<table>
<thead>
<tr>
<th>Row</th>
<th>Gene Name</th>
<th>Protein Name</th>
<th>Adj.P.Value</th>
<th>LogFC</th>
<th>Chr</th>
<th>K</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RHOA</td>
<td>ras homolog family member A</td>
<td>0.002</td>
<td>0.8</td>
<td>3</td>
<td>73</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>FGF2</td>
<td>fibroblast growth factor 2</td>
<td>0.030</td>
<td>0.9</td>
<td>4</td>
<td>76</td>
<td>0.02</td>
</tr>
<tr>
<td>3</td>
<td>FYN</td>
<td>FYN proto-oncogene, Src family tyrosine kinase</td>
<td>0.001</td>
<td>1.1</td>
<td>6</td>
<td>81</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>CD44</td>
<td>CD44 molecule (Indian blood group)</td>
<td>0.010</td>
<td>1.9</td>
<td>11</td>
<td>80</td>
<td>0.01</td>
</tr>
</tbody>
</table>
**Figure 3.** Action type analysis of the four significant central DEGs via CluePedia exploration is shown. Different colors indicate different action types; Green: Activation, Red: Inhibition, Purple: Catalysis, and Blue: Binding.

**Figure 4.** Biological process examination of the four significant expressed hub-bottlenecks in DS via ClueGO+ CluePedia. Four groups of biological processes are present here namely Negative regulation of blood vessel endothelial cell migration, Epiboly, hyaluronan catabolic process, and regulation of adherents junction organization, in which the most highlighted one is the first mentioned term in blue color. The four genes are assigned to different terms. The node colors indicate what terms are linked to that specific gene. For example, FYN only is related to one term, which is shown with that specific term.
Discussion

Down Syndrome as a genetic disorder is not well-studied in terms of molecular profiling (Liu et al., 2017). Identification of molecular events in DS could increase understanding of its mechanisms and consequently developing management approaches. One of which is examining differentially expressed genes in DS via array profiling. Furthermore, protein network mapping of these abnormal expressed genes could offer further understanding of the molecular behavior in an interacting system. For this aim, DS gene expression profile (GSE5390) was investigated via different online and offline analyzing tools. At first, the comparison of two groups of healthy and DS was conducted by box-plot through GEO2R online analyzer in figure 1. The comparison showed that the values are median-centered and consequently, the groups are comparable in terms of expression; hence, the data is suitable for further investigations. Analysis of GEO2R results showed that there are 63 up-regulated versus 3 down-regulated DEGs which differentiate DS group from healthy people. This fact is in correlation with previous reports that over-expression is dominant in chromosome 21 of Down syndrome brain (Liu et al., 2017; Lockstone et al., 2007; Mao, Zielke, Zielke, & Pevsner, 2003). Since PPI network analysis can be used to screen the top DEGs to find more effective and influence genes relative to onset and development of DS, the central elements (hub-bottlenecks) of network were identified and tabulated in Table S1. A total of 19 genes were assigned as hub-bottlenecks for the DS network in which, none of them were among top 250 DEGs except FYN. To evaluate rank of these central genes in DS expression profile, CluePedia was applied to combine the expression profile data with them as shown in figure 2. It can be inferred that all genes have expression values across the dataset GSE5390, except for MBOAT4. Since the DEGs which are expressed with at least one significant DE probe are considered as significantly expressed in DS, the genes of INS, EGF, ERBB2, AKT1, ALB, GAPDH, IL6, JUN, PIK3CA, PIK3CB, PRDM10, SRC, TP53, and CDC42 are not significantly expressed in DS. In another words, only four genes showed significant expression modifications among these 19 queried central genes. However, except RHOA which is with one spot, the other three DE hub-bottlenecks (FGF2, FYN, and CD44) in some spots show inhomogeneous expression patterns. These three genes significantly are
positively expressed in patients but not in some individual samples. Overall, these genes are considered statistically significant in expression since they have at least one spot with SDE.

The four significant central DEGs include RHOA, FGF2, FYN, and CD44 with 1, 2, 4, and 9 rows, respectively. All of which are up-regulated that may propose that no down-regulated gene may have noteworthy role in DS phenotype and provide more evidence for probable role of up-regulation in DS. Moreover, as indicated in table 1, none of these genes is located on chromosome 21; they are in fact on chromosome 3, 4, 6, and 11. This shows the effect of gene up-regulation event from other chromosomes in DS. Several experimental models in last decades have shown that small GTPases of the Rho family are master regulators of the actin cytoskeleton in every cell type (Hall, 2005). This is important to know that these molecules have critical linkage in several features of the neuronal differentiation (Govek, Newey, & Van Aelst, 2005) and key factors in many neurological syndromes and mental retardation (Newey, Velamoor, Govek, & Van Aelst, 2005; van Galen & Ramakers, 2005). Berto et al showed that the protein TTC3, encoded by one of the main down syndrome (Teeling et al.) critical region candidate genes, physically interacts with Citron kinase (CIT-K) and Citron N (CIT-N). These two are effectors of the RhoA small GTPase that have previously reported to be involved in neuronal proliferation and differentiation (Berto et al., 2007). As known enormous number of phenotypes, including learning difficulties, cardiac defects, unique facial features and leukemia accompanies with down syndrome (Lana-Elola, Watson-Scales, Fisher, & Tybulewicz, 2011). FGF2 cooperates stromal cell support of normal hematopoiesis by modulating osteoblast functions in bone marrow (Sugimoto et al., 2016). In which could have an association in leukemia of DS. Fyn is a tyrosine-specific phospho-transferase that is a member of the Src family of non-receptor tyrosine protein kinases (Resh, 1998). MM Ahmed et al considered the role of FYN in the nuclear fraction of hippocampus in DS models of rat (Ahmed et al., 2015). The CD44 is a cell surface marker, which is involved in cell–cell interactions, adhesion and migration. and encoded by the CD44 gene chromosome 11 (Spring et al., 1988). In addition, it is especially associates with active Src family protein tyrosine kinases Fyn in plasma membrane domains of human lymphocytes (Ilangumaran, Briol, & Hoessli, 1998). Moreover, as it is shown in table 1, FYN and RHOA are the most significantly expressed ones in DS. Additionally, CD44 showed the highest fold change among the DEGs. The analysis continued by examining the action type between these prominent genes (see figure 3). As it is apparent, all the queried action types are
present between FYN and RHOA. This phenomenon implies that these genes are in condense interactions. These two genes as indicated earlier are the most significant altered genes in DS as well. The other two genes, CD44 and FGF2 contributes in just two types of actions in this analysis, which is binding, and activation. To get a better view of the role of the important four genes, CluePedia handled the enrichment analysis. As it is presented in the figure 4, these genes contribute in four classes of 15 biological terms. The expression changes of our genes could have an impact on these biological processes and consequently, their dysregulation in DS. Negative regulation of blood vessel endothelial cell migration as the most highlighted group in our query shows that two genes of RHOA and FGF2 are involved in regulation of this class of 5 biological terms. CD44 and FYN are involved in two and one groups of biological processes, respectively. In addition, RHOA is almost participating in all groups except in hyaluronan catabolic process. Therefore, RHOA may have distinctive roles in DS due to its vast molecular characteristics.

Conclusion

It can be concluded that RHOA, FGF2, FYN, and CD44 (especially RHOA) and their related biological features may play indispensable associations in Down syndrome risk. Moreover, this study supports and suggests the fact that up-regulation may have more potential role in DS phenotype with a possible remarkable influence from other chromosome significantly differential expressed genes. However, more studies worth pursuing for verification of this finding.

Ethical Considerations

Compliance with ethical guidelines
The data were available as free access format from GEO database.

Funding
This study is related to the project NO.1396/46460 from Student Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran. This study was supported by “Student Research Committee “ and “Research & Technology Chancellor “ in Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Conflict of interest
There is no conflict of interest

Acknowledgements
We also appreciate the “Student Research Committee” and “Research & Technology Chancellor” in Shahid Beheshti University of Medical Sciences for their financial support of this study.

References


Spring, F., Dalchau, R., Daniels, G., Mallinson, G., Judson, P., Parsons, S., . . . Anstee, D. (1988). The Ina and Inb blood group antigens are located on a glycoprotein of 80,000 MW (the CDw44 glycoprotein) whose expression is influenced by the In (Lu) gene. *Immunology, 64*(1), 37-43.


