

# A Comprehensive Bioinformatics Approach to Identify Molecular Signatures and Key Pathways for the Huntington Disease

Bioinformatics and Biology Insights  
Volume 17: 1–17  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11779322231210098



Tahera Mahnaz Meem<sup>1</sup>, Umama Khan<sup>2</sup>, Md Bazlur Rahman Mredul<sup>1</sup>,  
Md Abdul Awal<sup>3</sup>, Md Habibur Rahman<sup>4</sup> and Md Salauddin Khan<sup>1</sup>

<sup>1</sup>Statistics Discipline, Science, Engineering & Technology School, Khulna University, Khulna, Bangladesh. <sup>2</sup>Biotechnology & Genetic Engineering Discipline, Khulna University, Khulna, Bangladesh. <sup>3</sup>Electronics and Communication Engineering Discipline, Khulna University, Khulna, Bangladesh. <sup>4</sup>Department of Computer Science and Engineering, Islamic University, Kushtia, Bangladesh.

**ABSTRACT:** Huntington disease (HD) is a degenerative brain disease caused by the expansion of CAG (cytosine-adenine-guanine) repeats, which is inherited as a dominant trait and progressively worsens over time possessing threat. Although HD is monogenetic, the specific pathophysiology and biomarkers are yet unknown specifically, also, complex to diagnose at an early stage, and identification is restricted in accuracy and precision. This study combined bioinformatics analysis and network-based system biology approaches to discover the biomarker, pathways, and drug targets related to molecular mechanism of HD etiology. The gene expression profile data sets GSE64810 and GSE95343 were analyzed to predict the molecular markers in HD where 162 mutual differentially expressed genes (DEGs) were detected. Ten hub genes among them (*DUSP1*, *NKX2-5*, *GLI1*, *KLF4*, *SCNN1B*, *NPHS1*, *SGK2*, *PITX2*, *S100A4*, and *MSX1*) were identified from protein-protein interaction (PPI) network which were mostly expressed as down-regulated. Following that, transcription factors (TFs)-DEGs interactions (FOXC1, GATA2, etc), TF-microRNA (miRNA) interactions (hsa-miR-340, hsa-miR-34a, etc), protein-drug interactions, and disorders associated with DEGs were predicted. Furthermore, we used gene set enrichment analysis (GSEA) to emphasize relevant gene ontology terms (eg, TF activity, sequence-specific DNA binding) linked to DEGs in HD. Disease interactions revealed the diseases that are linked to HD, and the prospective small drug molecules like cytarabine and arsenite was predicted against HD. This study reveals molecular biomarkers at the RNA and protein levels that may be beneficial to improve the understanding of molecular mechanisms, early diagnosis, as well as prospective pharmacologic targets for designing beneficial HD treatment.

**KEYWORDS:** Huntington disease, bioinformatics, system biology, biomarkers, hub genes, pathway

**RECEIVED:** January 6, 2023. **ACCEPTED:** October 7, 2023.

**TYPE:** Original Research Article

**FUNDING:** The author(s) received no financial support for the research, authorship, and/or publication of this article.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHORS:** Md Habibur Rahman, Department of Computer Science and Engineering, Islamic University, Kushtia 7003, Bangladesh. Email: habib@iu.ac.bd

Md Salauddin Khan, Statistics Discipline, Science, Engineering and Technology School, Khulna University, Khulna-9208, Bangladesh. Email: salauddin@stat.ku.ac.bd

## Introduction

Huntington disease (HD) can be defined as a completely penetrant degenerative brain disease resulting from CAG (cytosine-adenine-guanine) repeat expansion that is inherited as a dominant trait in IT15 (“interesting transcript 15”) gene also defined as huntingtin (*HTT*) gene.<sup>1,2</sup> Huntingtin is translated into such an excessively massive polyglutamine domain around the N-terminus of the Huntington protein.<sup>3</sup> Mutant huntingtin (mHTT) becomes particularly unstable due to the enlargement of the CAG region and allows to a combination of proteins of the same type and/or other types resulting in clumps that may lead to discontinued neurotransmission.<sup>4,5</sup>

Huntington disease is caused by a complicated interaction of environmental and genetic factors, resulting in succeeding generations to live in increasingly deplorable social conditions. Autosomal recessive inheritance is often seen in mid-life diagnosis and

continuous progressive motor including intellectual and mental symptoms across 15 to 20 years.<sup>6,7</sup> It commonly starts in middle age, between the 21 and 50 years, including an average starting age of 41 years.<sup>6,8</sup> Fine motor function deterioration, cerebellar abnormalities, gait abnormalities, dysarthria, cognitive difficulties, and stiffness are all prevalent symptoms and indicators.<sup>9</sup> It produces substantial physical and cognitive deficiencies such as memory loss, depression, mood fluctuations clumsiness as well as some other psychological problems and disorders.<sup>4,10</sup> Clinical diagnosis to identify HD is conducted when there are strong sets of evidence of a motor disorder, particularly chorea involving iatrogenic conditions and general internal disorders.<sup>9</sup> Degenerative mutations and cellular damage can be found throughout the cortex and external areas of the central nervous system, specifically across the striatum; there are various symptomatic therapies and medication available, but there is no permanent remedy for this devastating brain disorder presently.<sup>10,11</sup> There are currently no medications available to delay symptoms and disease development; however, there are numerous effective post management (ie, medication and nonpharmacologic approaches available).<sup>12</sup>

Umama Khan is now affiliated to Department of Genetic Engineering and Biotechnology, Faculty of Biological Science and Technology, Jashore University of Science and Technology, Jashore, Bangladesh



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

As the disease was first diagnosed, especially after Davenport and Munceyat 1916 pointed out a large collection of information on households in the New England area of America, individual family surveys on HD have been conducted.<sup>13</sup> Most European nations, including Northern and Southern, have a quite high prevalence, ie, 4 to 8 in every 100 000 individuals, and the disease may also be common in India side by side in some areas of Central Asia, according to the findings.<sup>14</sup> The prevalence of HD worldwide appears to vary by region.<sup>15</sup> In South Africa and Zimbabwe, prevalence rates were determined to be extremely low.<sup>16</sup> Since 1995, the prevalence rate (average) in Asian countries (South Korea, Hong Kong, Taiwan, and Japan) has been low (0.42 per 100 000 individuals) compared with the rate among white populations (North America, Australia, the United Kingdom and Western Europe) where the rate was 9.71 per 100 000 individuals.<sup>15</sup> Following the emergence of genetic testing, the frequency of HD has increased in several populations.<sup>17</sup>

Although several studies have already identified histone modifications, protein hubs identification, transcription factor (TF) impairments, and abnormal microRNA (miRNA) expression, HD is difficult to diagnose at an early stage and identification is restricted in accuracy and precision, as well as expensive.<sup>18,19</sup> Also, contradictions between researchers were found about the conclusion of differentially expressed genes. As a result, using brain cell analysis to predict HD could enhance disease's early detection and treatment. In 1993, the faulty gene that triggers HD was first discovered and a genetic test for diagnosis is available now.<sup>20</sup> The test can identify the malfunctioning gene for HTT protein as well as discover the genetic defect in individuals who already haven't shown any symptoms but has the possibility to obtain the disease but this is a slow process and can only be diagnosed after mid-age.<sup>20,21</sup> This condition is caused by a CAG triplet repeat amplification in the *HTT* gene.<sup>20,22</sup> Due to prolonged CAG repeat, HTT protein with an enlarged polyglutamine tract is produced, resulting in pathogenic HTT protein residues that are immune to the protein cycle, leading to cellular toxicity.<sup>23,24</sup> Most HD patients have motor-involved problems, a small percentage (15%) establish clinically relevant psychological disease first and hence have a mental diagnosis before the start of movement disorder.<sup>20,25</sup> Huntington disease is associated with extremely selective deterioration of the corpus striatum which is a brain region with no noticeable abnormalities in peripheral tissues. Huntington and mHTT are prevalently distributed across the brain also including peripheral tissues.<sup>10</sup>

Histone modifications, protein hubs identification, TF impairments, and abnormal miRNA expression are among the processes linked to this imbalance and identification of disease specification biomarkers is crucial for HD drug testing.<sup>26</sup> Several studies have been conducted from different perspectives to identify molecular biomarkers with various brain and blood data. As HD is a completely penetrant degenerative brain disease, in this study brain-based data sets were obtained to progress further analyses. Generally, genetic factors in HD have primarily been examined in diseased brains comparing controlled individuals, as well as in cell and animal models.<sup>23,27</sup> Hundreds of research

teams have spent the past 3 decades trying to figure out what causes HD based on cellular and molecular compounds.<sup>18,19,27</sup> The goal of this research is to find molecular biomarkers that represent the brain expression alignments of associative factors connected to HD development. The main objective of this study is to identify the differentially expressed genes (DEGs) and common genes using genomic information for HD. Second, we aimed to identify molecular biomarkers to diagnose early detection of HD and progression by identifying transcriptional and post-transcriptional level; side by side another objective was to identify potential interacting components that are prominent in the gene list which can be used to find valuable medicines or drug targets. The rest of this article is organized as materials and methodology, results, discussion, and conclusions section. The materials and methodology section contains data acquisition, determination of DEGs and common genes, gene set enrichment analysis (GSEA), construction of protein-protein interaction (PPI) network for hub protein identification, transcriptomic marker identification, prediction of small drug molecules, and finally, validation of the predicted molecular markers.

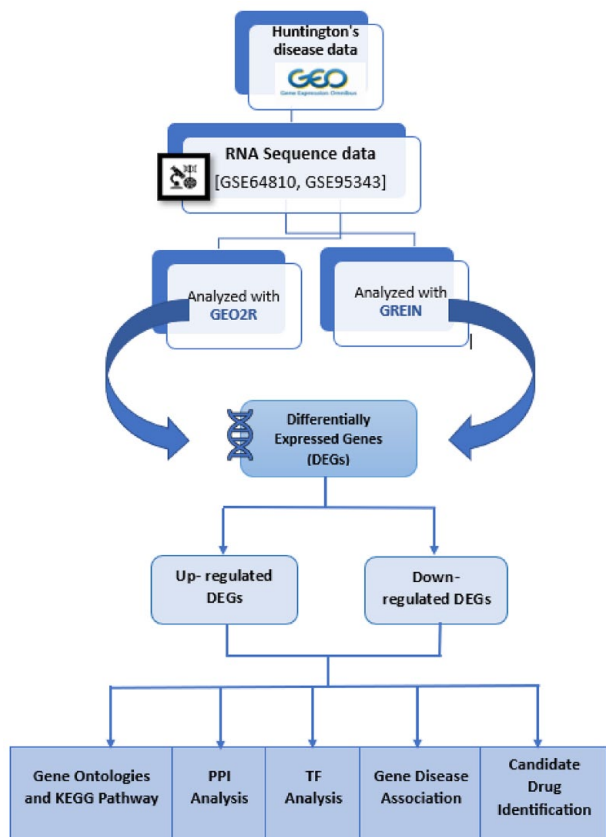
## Materials and Methodology

### *Data acquisition*

In this study, data sets were obtained from National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO).<sup>28</sup> We searched for *Homo sapiens* data of HD and detected 747 data sets, the majority of which were initially disqualified as they were noncoding data sets, had a very small sample size, were redundant data sets, had an inadequate format, or inappropriate experimental set-up, contained stem-cells, did not contain control subjects, or was from organisms other than humans. Finally, after considering all perspectives, we have selected 2 data sets for the analysis which are GSE64810 and GSE95343, where GSE64810 contained mRNA sequence expression of brain tissue and GSE95343 contained mRNA sequence of neural cells.

### *Determination of differentially expressed genes and common genes*

Gene set enrichment analysis identifies DEGs from a large set of genes linked to disease symptoms using a number of statistical approaches.<sup>29,30</sup> Figure 1 visualizes the steps involved in data acquisition and identifying DEGs also shows the analytical approaches in a nutshell. GSE64810 and GSE95343 shared a total of 162 common genes, whereas 106 genes were up-regulated and 56 genes were down-regulated. Gene expressions of the data sets were analyzed in accordance with  $P$  value  $< .05$  and  $\log_2$  absolute values for Fold Control (FC); ( $-1 \geq FC \geq 1$ ). Linear models identified the DEGs where (limmapackage obtained from Bioconductor using R) R or the online tool GEO2R for microarray data sets and GREIN<sup>31</sup> for RNA sequence data sets were used. The limma is known as a Bioconductor software package for analyzing data using R from gene expression studies and experiments.<sup>32</sup>  $P$  value was



**Figure 1.** Systematical implementation approach in the study: significant differentially expressed genes were identified as well as associated gene ontology terminologies and pathways were enriched using mRNA expression data sets of Huntington disease from neural cells and brain tissue. Multiple network approaches were also implemented to identify PPI, regulatory signature molecules, and potential therapeutic drug targets. DEGs indicate differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein-protein interaction; TF, transcription factor.

adjusted using the Benjamini-Hochberg (BH) technique where the FDR (false discovery rate),  $Q_e$  is defined as the expectation of  $Q^{33}$  and it implies that

$$Q_e = E(Q) \quad (1)$$

The random variable indicates the fraction of errors made by the null hypotheses that are falsely rejected and

$$Q = \frac{V}{V + S} \quad (2)$$

$$R = V + S \quad (3)$$

where

$V$  = number of significant true null hypotheses,

$S$  = number of significant nontrue null hypotheses, and

$R$  = number of hypotheses that are rejected.

Finally, it can be seen that

$$E(Q) = E\left(\frac{V}{R}\right) \quad (4)$$

Common genes were obtained from the DEGs using the Venn diagram comparing the common DEGs.

### Gene set enrichment analysis

Gene set enrichment analysis was conducted to depart analytical obstructions and examine data if individuals from a set of genes tend to appear near the highest points (or bottom) of the list and also checks where the gene set is linked to behavioral class distinction.<sup>34,35</sup> To obtain GSEA, gene ontologies (molecular function, biological process as well as cellular components), molecular pathways (Kyoto Encyclopedia of Genes and Genomes [KEGG] pathway) and was obtained using Enricher, GREIN, David Bioinformatics Resources 6.8 and R programming, respectively.<sup>36,37</sup> A considerable analytical effort is required to identify common biological findings using GSEA. An adjusted  $P < .05$  was used as the cut-off parameter to conduct gene enrichment. The KEGG pathways were performed by David Bioinformatics Resources 6.8 and the significant KEGG pathway was obtained underlying the  $P < .05$ . A bubble plot was also generated using the R tool ggplot2 to visualize KEGG pathways where  $P$  value ( $P < .05$ ) and the FDR value ( $FDR < .05$ ) was demarked. It is commonly used to comprehend metabolic pathways to annotate genes.

### Determination of protein-protein interaction

Protein-protein interaction (PPI) is the protein interactions with other proteins, ions, and nonprotein elements like lipids, carbohydrates, and nucleic acids generate the distinctive features of each protein.<sup>38</sup> Protein-protein interaction plays an important role in several biological functions and includes a variety of applications in biology regarding structure, affinities, and whether or not the interaction is persistent.<sup>39</sup> Interaction of protein hubs through PPI networks was retrieved using Network Analyst<sup>40</sup> from the STRING database depending on the physical connection of common gene proteins selecting confidence score 900 (adjusted  $P < .05$ ). We used Cytoscape software<sup>41</sup> including cytoHubba and MCODE applications<sup>40</sup> to identify hub genes.

### Analysis of transcription factor gene and transcription factor-microRNA coregulatory interconnections

Transcription factors interact directly with DNA to bind the regulatory regions known as regulators and enhancers, which increase (or occasionally prevent) gene transcription and consequently messenger RNA synthesis.<sup>42,43</sup> Network Analyst was used to determine regulatory TFs-DEGs interaction network and also to identify TFs and miRNAs that control common genes at the transcriptional and post-transcriptional stages. The TF-genes and TF-miRNA were obtained from the JASPAR TF binding site profile database and TarBase v8.0, respectively. Here, JASPAR is a popular manually organized, nonredundant open-accessed TF binding profiling database recorded as matrices and models of TF.<sup>44</sup> TarBase is a comprehensive database of miRNA-gene interactions that had been confirmed empirically.<sup>45</sup>



**Table 1.** Statistical overview for gene expression in data sets used in analysis.

SERIAL NUMBER	GEO ACCESSION	GEO PLATFORM	SOURCE	NUMBER OF DEGS			COMMON GENES
				UP	DOWN	TOTAL	
1	GSE64810	GPL11154	Prefrontal cortex (post-mortem BA9 brain tissue)	715	521	1236	162
2	GSE95343	GPL11154	Neural cells	1297	1799	3096	
Total				2012	2320	4332	

Abbreviations: DEGs, differentially expressed genes; GEO, Gene Expression Omnibus.

### Gene disease association

Genetic and/or environmental factors are thought to play a role in complex disorders.<sup>46</sup> Recognizing the genes linked to an illness can help with disease treatment, prevention, and health care.<sup>47</sup> It is also crucial for knowing how genes work biologically. It emphasizes the evolution of human genetic illnesses which are used to uncover linkages between hereditary and genetic disease while using Network Analyst to investigate related diseases and common DEGs. The DisGeNET database, which is exclusively suitable for human data sets, was used to gather the gene-disease association information. DisGeNET is a research platform that includes one of the most comprehensive publicly available libraries of genes and variations linked to human disorders.<sup>48</sup>

### Prediction of candidate drugs or small compounds

Enrichment of drug-target was executed using Enrichr to obtain candidate drugs or small compounds. This method was used to find small chemical that influenced HD function. The gene sets (DSigDB) database from the Enrichr was employed to identify prospective drugs or small molecules. The Drug Signatures Database (DSigDB) is a gene set database that connects drugs/compounds with respective target genes.<sup>49</sup> Chemical structures of the candidate drugs were extracted from PubChem.<sup>50</sup>

### Validation of relevant biomarkers

To check the validation of relevant biomarkers revealed in this research, we conducted an overall review from literatures with respect to the findings of our study.<sup>51,52</sup>

## Results

### Determination of differentially expressed genes and overlapping genes

High-throughput sequencing data (GSE64810, GSE95343) of prefrontal cortex and neural cells of HD were analyzed using GEO2R and GREIN. GSE64810 data were obtained by expression analysis of human post-mortem brain tissue (BA9) for HD patients and neurologically normal people

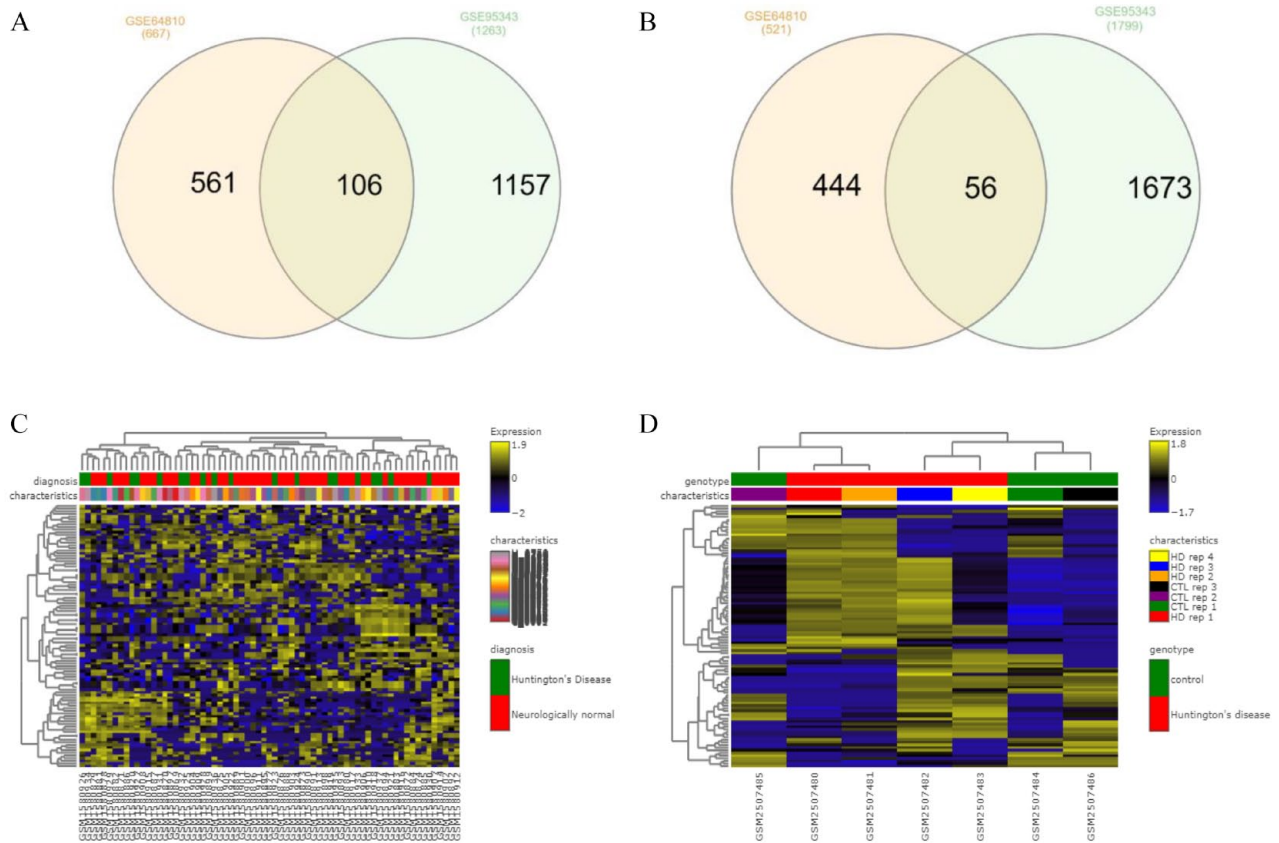
using mRNA-Seq. GSE95343 data set consists of neural cells for HD and control patient using mRNA-Seq. The transcriptomic HD data sets exhibited 162 common DEGs including up-regulated and down-regulated genes from overlapped DEGs (Table 1 and Figure 2). Figure 2A and B depicts interactive heatmaps for control and case samples from the selected data sets.

### Gene set enrichment analysis

Significant Gene Ontology (GO) was enriched through molecular function, biological process as well as cellular components (Table 2). Significant GO terms were displayed in Table 2 and Figure 3C to E, which was obtained using David Bioinformatics Resources 6.8. In Biological process, 13 genes were involved in proteolysis and 8 genes were found to be engaged in transcription from skeletal system development. The cellular component visualized 46 genes that came from extracellular region, whereas extracellular space (40 genes) and extracellular exosome (32 genes) also involved majority numbers of genes. Molecular function of the GO expressed 18 genes (EOMES, OSR2, PRRX2, etc) which were involved in sequence-specific double-stranded DNA binding. Figure 3A illustrates the DEG-enriched molecular pathways that have been found in KEGG Pathway with help of Enrichr and Figure 3B represents the bubble plot following KEGG Pathway produced by R programming. In the KEGG pathway, antigen processing and presentation pathway was seen to be enriched with 6 genes significantly, according to significant *P* value. Top significantly enriched pathways were represented by bubble plots in this study.

### Protein-protein interaction

Figure 4 visualizes the PPI obtained by network analysis highlighting the protein hubs with the list of top 10 hub genes produced by Cytoscape, and it shows the interactions between the hub genes generated by Network Analyst and Cytoscape. Protein-protein interaction network of the proteins was analyzed to identify the protein hubs of respective DEGs including 162 common genes. Network analysis determined the hub proteins, and they were *DUSP1*, *NKX2-5*,



**Figure 2.** (A) Overlapping 106 up-regulated common genes among data sets GSE64810 and GSE95343; (B) Overlapping 56 down-regulated common genes among data set GSE64810 and GSE95343; (C and D) heatmap of selected data sets, GSE64810 and GSE95343 displaying differential expression level; the color scale reflects the expression value, and each row and column represents a gene. DEGs indicates differentially expressed genes.

*GLI1*, *KLF4*, *SCNN1B*, *NPHS1*, *SGK2*, *PITX2*, *S100A4*, and *MSX1*. These hub genes could be the indicators that play a key role in the progression of HD.

#### Transcription factor gene and transcription factor-microRNA coregulatory interconnections

To get insights into substantial changes happening at the transcriptional and post-transcriptional level, the DEGs were screened to pick out the transcriptional (TFs) and post-transcriptional (miRNAs) regulatory biomolecules which were identified based on degree value. We predicted regulatory biomolecules that may influence DEGs at the transcription regulatory and post-transcriptional stages by identifying TFs and miRNAs targeting DEGs. Figure 5A and B exhibits the analysis of TF-gene and TF-miRNA coregulatory interconnections with network. The analysis disclosed top significant transcriptomic factors (TFs), ie, *FOXC1*, *GATA2*, *YY1*, *TFAP2A*, *FOXL1*, *PPARG*, *HINFP*, *STAT3*, *MEF2A* and top significant miRNAs, ie, *hsa-miR-340*, *hsa-miR-34a*, *hsa-miR-495*, *hsa-miR-1*, *hsa-miR-124*, *hsa-miR-29a*, *hsa-miR-29b*, *hsa-miR-30e*, *hsa-miR-16*, *hsa-miR-206*, *hsa-miR-30a*, *hsa-miR-30c*, and *hsa-miR-944* as regulatory biomarkers for HD.

#### Gene disease association

Figure 6 indicates the main diseases linked to common DEGs are highlighted in the gene-disease association network. Degenerative polyarthritis, schizophrenia, short stature, autosomal recessive predisposition, and liver cirrhosis (Experimental) were identified as the correlated diseases to the DEGs from gene-disease association study using Network Analyst.

#### Development of candidate drugs or small compounds

Table 3 lists the top associated small molecules, along with respective molecular formula and structural composition, and Figure 7 shows a bar graph of the small drug molecules produced by Enrichr. Cytarabine exhibited the strongest negative association and the best chance of reversing HD among these compounds. Arsenite has also shown effectiveness against HD.

#### Validation of relevant biomarkers

The hub genes we identified were *DUSP1*, *NKX2-5*, *GLI1*, *KLF4*, *SCNN1B*, *NPHS1*, *SGK2*, *PITX2*, *S100A4*, and *MSX1* reflecting as the potential biomarkers for HD. To determine

**Table 2.** Functional enrichment approach of DEGs to find Gene Ontology entries for the HD based on top significant terms.

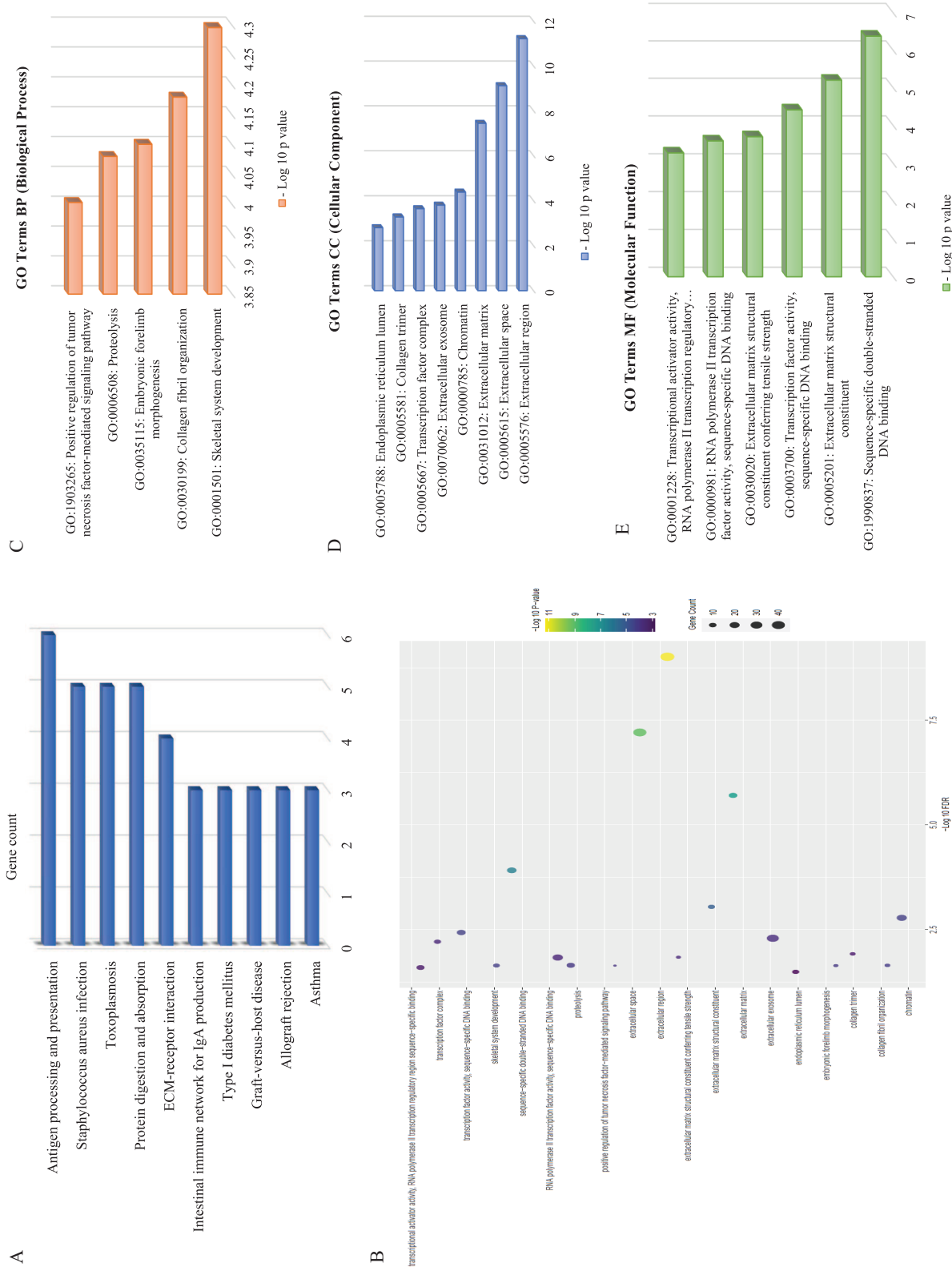
CATEGORY	GO ID	TERM	GENE COUNT	P (< 0.5)	GENES	FDR (< 0.5)
Biological process	GO:0001501	Skeletal system development	8	$5.05 \times 10^{-05}$	COL1A1, COL3A1, COL1A2, FRZB, ALX4, TNFRSF11B, HOXA13, BMP5	.022761
	GO:0030199	Collagen fibril organization	6	$6.61 \times 10^{-05}$	COL1A1, COL3A1, ADAMTS2, SFRP2, COL1A2, CYP1B1	.022761
	GO:0035115	Embryonic forelimb morphogenesis	5	$7.93 \times 10^{-05}$	OSR2, CRABP2, ALX4, MSX1, HOXA13	.022761
	GO:0006508	Proteolysis	13	$8.31 \times 10^{-05}$	CFH, FGL2, PCOLCE, CPZ, CTSS, TMPRSS15, FOLH1, ANPEP, CASP4, CASP1, REN, CPVL, CFB	.022761
	GO:1903265	Positive regulation of tumor necrosis factor-mediated signaling pathway	4	$9.92 \times 10^{-05}$	CASP4, CASP1, HSPA1B, HSPA1A	.022761
Cellular component	GO:0005576	Extracellular region	46	$6.02 \times 10^{-12}$	CXCL6, PTGFR, ECM2, CFH, FGL2, TNFRSF11B, OLFML2A, NID1, CTSS, FGF6, WNT6, HPX, ADAMTS2, IGFBP1, FRZB, PDGFD, CASP4, OLFML1, CASP1, S100A11, TSLP, OMD, APOA2, KERA, PCOLCE, C4ORF48, DCN, PDGFRL, BMP5, TGFBFR3, MFAP5, COL1A1, COL3A1, SFRP2, COL1A2, COL6A2, MFAP2, MGP, OGN, COL6A3, S100A4, REN, AVP, HSPA1B, CFB, HSPA1A	$9.75 \times 10^{-10}$
	GO:0005615	Extracellular space	40	$7.59 \times 10^{-10}$	CXCL6, SERPINB11, CSTA, CFH, ASIP, FGL2, TNFRSF11B, SOGA3, OLFML2A, CPZ, CTSS, FGF6, WNT6, HPX, IGFBP1, FRZB, ANPEP, PDGFD, OLFML1, S100A11, TSLP, OMD, KERA, PCOLCE, DCN, BMP5, TGFBFR3, COL1A1, COL3A1, SFRP2, COL1A2, COL6A2, OGN, COL6A3, S100A4, REN, FAM20A, AVP, CFB, CCL26	$6.15 \times 10^{-08}$
	GO:0031012	Extracellular matrix	14	$3.64 \times 10^{-08}$	ECM2, KERA, TNFRSF11B, OLFML2A, DCN, TGFBFR3, COL1A1, WNT6, COL3A1, ADAMTS2, COL1A2, COL6A2, MGP, COL6A3	$1.96 \times 10^{-06}$
	GO:0000785	Chromatin	21	$4.14 \times 10^{-05}$	EOMES, BNC2, PRRX2, SIX1, BARY1, KLF4, MEOX2, HOXA13, TBX18, POU5F1, NPAS4, TBX15, LHX1, SIX2, ALX4, OTX2, MSX1, NKX2-5, SIM1, PITX2, NEUROG2	.001678
	GO:0070062	Extracellular exosome	32	$1.58 \times 10^{-04}$	SLC26A2, PTPRS, CFH, CRABP2, FGL2, NID1, WNT6, HPX, ANPEP, S100A11, LRRC26, OMD, SLC6A13, APOA2, PCOLCE, TGFBFR3, FABP1, FOLH1, COL1A2, NPHS1, SCNN1B, COL6A2, MGP, OGN, COL6A3, HLA-DRA, S100A4, FAM20A, CPVL, HSPA1B, CFB, HSPA1A	.005124

(Continued)

Table 2. (Continued)

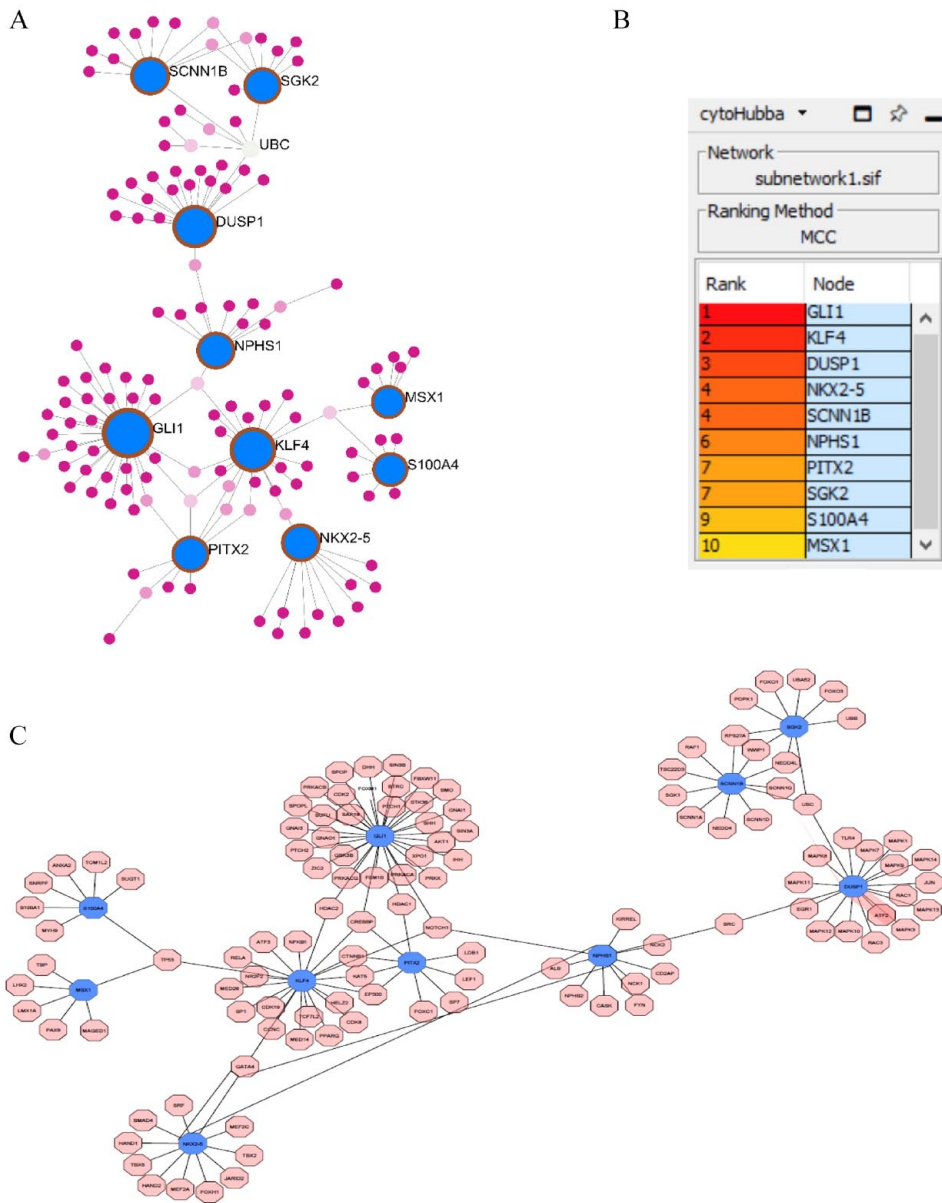
CATEGORY	GO ID	TERM	GENE COUNT	P (< 0.5)	GENES	FDR (< 0.5)
	GO:0005667	Transcription factor complex	9	$2.27 \times 10^{-04}$	NPAS4, SIX2, LHX1, ALX4, SIX1, KLF4, NKX2-5, PITX2, POU5F1	.00613
	GO:0005581	Collagen trimer	6	$5.27 \times 10^{-04}$	COL1A1, COL3A1, COL1A2, COL6A2, COL6A3, PCOLCE	.012195
	GO:0005788	Endoplasmic reticulum lumen	9	.001589	COL1A1, WNT6, COL3A1, COL1A2, PDGFD, COL6A2, APOA2, GPX8, COL6A3	.032174
Molecular function	GO:1990837	Sequence-specific double-stranded DNA binding	18	$4.20 \times 10^{-07}$	EOMES, OSR2, PRRX2, SIX1, KLF4, MEOX2, HOXA13, TBX18, POU5F1, TBX15, LHX1, SIX2, ALX4, OTX2, MSX1, NKX2-5, PITX2, NEUROG2	$1.22 \times 10^{-04}$
	GO:0005201	Extracellular matrix structural constituent	9	$6.32 \times 10^{-06}$	MFAP5, COL1A1, COL3A1, COL1A2, MGP, FGL2, MFAP2, PCOLCE, NID1	$9.16 \times 10^{-04}$
	GO:0003700	Transcription factor activity, sequence-specific DNA binding	15	$3.88 \times 10^{-05}$	EOMES, PRRX2, SIX1, BARX1, KLF4, MEOX2, TBX18, POU5F1, TBX15, LHX1, SIX2, OTX2, NKX2-5, SIM1, PITX2	.003754
	GO:0030020	Extracellular matrix structural constituent conferring tensile strength	5	$1.99 \times 10^{-04}$	COL1A1, COL3A1, COL1A2, COL6A2, COL6A3	.01445
	GO:0000981	RNA polymerase II transcription factor activity, sequence-specific DNA binding	22	$2.56 \times 10^{-04}$	EOMES, OSR2, PRRX2, SIX1, BARX1, KLF4, GLI1, MEOX2, HOXA13, TBX18, POU5F1, NPAS4, TBX15, LHX1, SIX2, ALX4, OTX2, MSX1, NKX2-5, SIM1, PITX2, NEUROG2	.014871
	GO:0001228	Transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	12	$5.35 \times 10^{-04}$	NPAS4, OSR2, SIX2, ALX4, SIX1, OTX2, MSX1, KLF4, NKX2-5, MEOX2, GLI1, HOXA13	.025857

Abbreviations: DEG, differentially expressed gene; FDR, false discovery rate; GO, Gene Ontology; HD, Huntington disease.



**Figure 3.** (A) KEGG Pathway enriched with the relevant DEGs,  $P < .05$ ; Network Analyst and the Enrichr web-based server identified KEGG pathways ( $P < .05$ ) linked with substantial DEGs. (B) DEG-related pathways in KEGG pathway analysis; the color of the bubble figure denotes  $\log_{10} P$  values, whereas the size of the bubbles reflects gene counts. Furthermore, the horizontal axis reflects  $\log_{10}$  FDR. (C to E) The bar diagrams illustrate significant GO terms ( $P < .05$ ) related with genes regarding biological process, cellular component, and molecular function categories, correspondingly. The vertical axis displays enhanced GO terms with respect to ( $P < .05$ ) and horizontal axis displays  $-\log_{10}(P$  values). DEGs indicates differentially expressed genes; ECM, extracellular matrix; FDR, false discovery rate; GO, Gene Ontology; IgA, immunoglobulin A; KEGG, Kyoto Encyclopedia of Genes and Genomes.

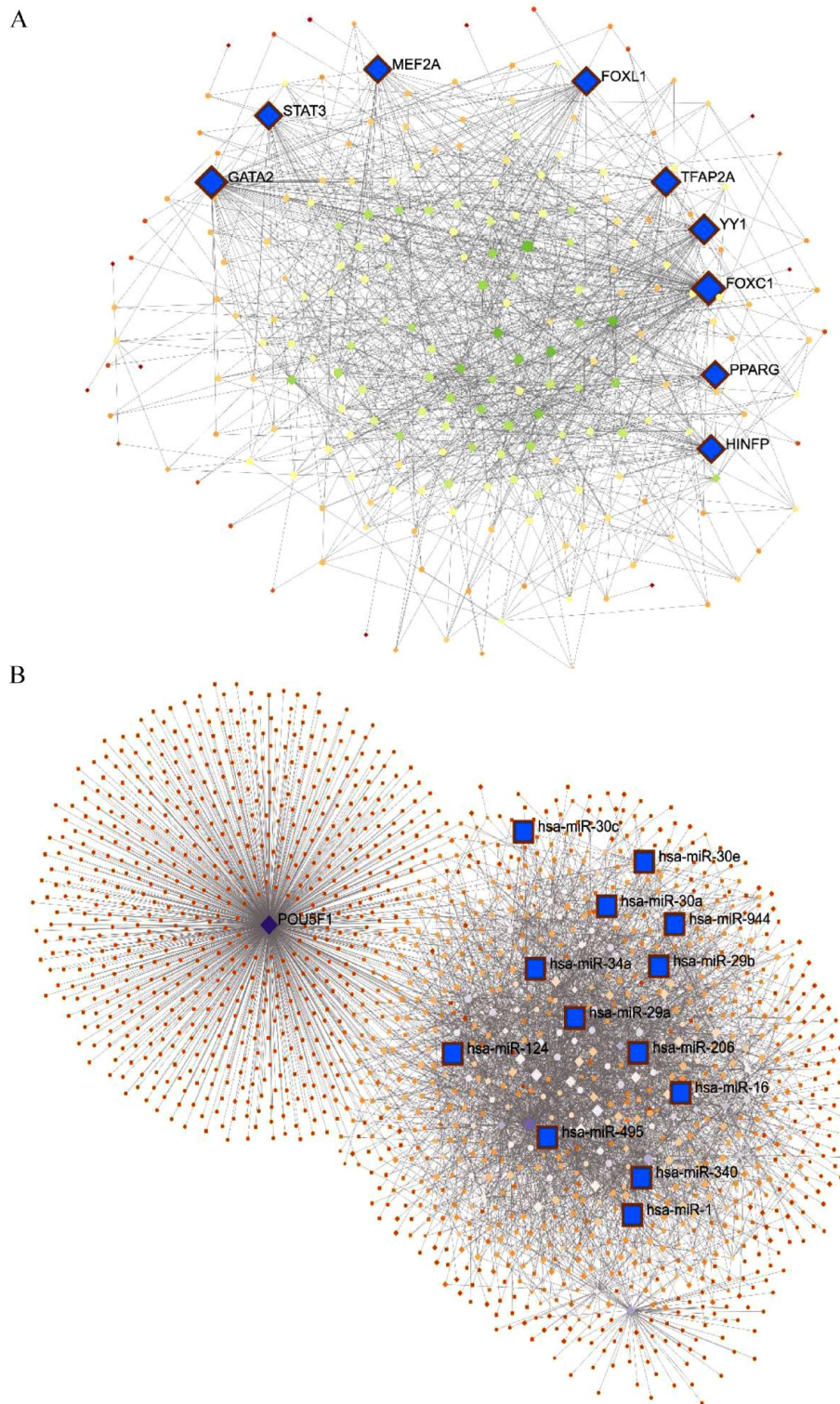




**Figure 4.** Visualization of PPI network of common DEG: (A) PPI network developed by Network Analyst showing the top 10 hub genes that interconnect other DEGs marked by blue nodes and other nodes symbolize DEGs connected by edges which reflect DEG interaction. (B) List of top 10 hub genes produced by Cytoscape through MCC method. Here, color shades represent the rank of hub proteins. The darker shades of color express the higher rank of hub genes and vice versa. (C) According to the selected layout, the top 10 hub proteins in a network are visualized using the MCC technique in Cytoscape v3.8.2 using the cytoHubba plugin. Pink-colored nodes represent the proteins associated via edges with the top 10 hub proteins in blue shade. DEG indicates differentially expressed gene; MCC, maximum clique centrality; PPI, protein-protein interaction.

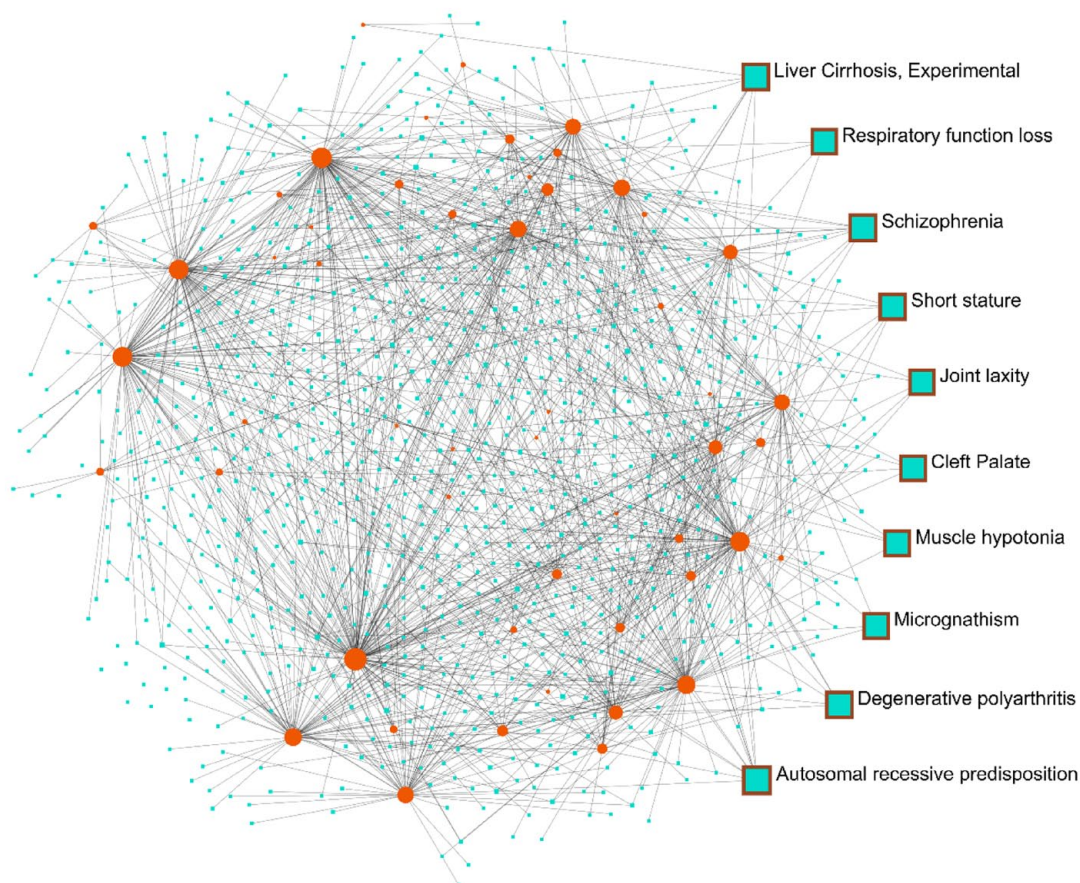
the validity of these putative biomarkers, we looked through the related studies for prospective HD biomarkers (Figure 8). It was identified that low DUSP1 also known as MKP-1 expression may contribute to hyperactivation of MAPK in HD, whereas increased MKP activity may be neuroprotective.<sup>53-55</sup> NKX2-5 was found to be correlated with the protein synthesis in HD<sup>56</sup> and also as one of the significantly up-regulated genes.<sup>57</sup> GLI1 was also found as contributing gene of HD, and we found that it was one of the hub proteins that contribute to HD.<sup>58-63</sup> Previous studies suggested KLF4 as a

potential biomarker for HD which is consistent with our study.<sup>64-68</sup> SCNN1B's utility as a molecular biomarker in HD was demonstrated in 2 earlier studies.<sup>69,70</sup> Increased expression of NPHS1 has consistently been identified to HD, which supports our research results.<sup>71</sup> In our research, it was revealed that SGK2 is a possible protein hub in HD, and it has been seen to be widely expressed in a study published in 2011.<sup>72</sup> Over-expressed PITX2 have effects on the initial formation of striatal neurons composing the pathways directly and indirectly, according to few studies. In our study, PITX2 was recognized



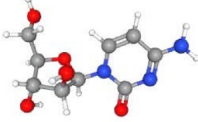
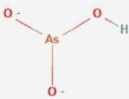
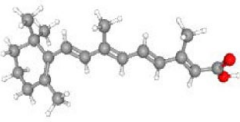
**Figure 5.** TF-gene and TF-miRNA coregulatory interconnections. (A) Identification of transcriptomic regulatory signatures by network analysis of DEGs using the Network Analyst server: the network highlights the top 9 TFs that are linked to DEGs. Blue nodes represent the TFs, other nodes represent the DEGs, and edges represent the interaction levels. (B) Network-based analysis of transcriptomic regulatory signatures by network analysis of DEGs using the Network Analyst server: the network displays the top 12 miRNAs interconnected with DEGs, with blue nodes representing miRNAs and other nodes representing DEGs linked through edges. DEGs indicates differentially expressed genes; TF, transcription factor.





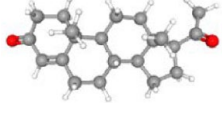
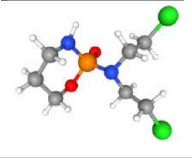
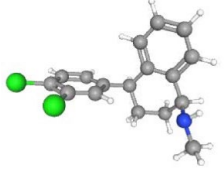
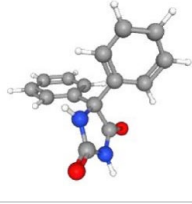

**Figure 6.** The main diseases linked to common differentially expressed genes are highlighted in the gene-disease association network (green: seeds and blue: diseases).

**Table 3.** List of probable Huntington disease medications, along with respective chemical formula and structural composition.

TERM	<i>P</i>	CHEMICAL FORMULA	GENES	STRUCTURAL COMPOSITION
Cytarabine CTD 00005743	3.64E-07	$C_9H_{13}N_3O_5$	<i>BNC2</i> ; <i>CPZ</i> ; <i>DCN</i> ; <i>SLC7A2</i> ; <i>BMP5</i> , etc	
Arsenite CTD 00000779	5.76E-07	$AsO_3^{3-}$	<i>CFH</i> ; <i>SIX1</i> ; <i>HTR2C</i> ; <i>SOGA3</i> ; <i>BARX1</i> , etc	
Methotrexate CTD 00006299	2.91E-06	$C_{20}H_{22}N_8O_5$	<i>SLC26A2</i> ; <i>CSTA</i> ; <i>DUSP1</i> ; <i>TNFRSF11B</i> ; <i>CTSS</i> , etc	
Retinoic acid CTD 00006918	3.05E-06	$C_{20}H_{28}O_2$	<i>PTPRS</i> ; <i>BNC2</i> ; <i>SIX1</i> ; <i>HTR2C</i> ; <i>SOGA3</i> , etc	

(Continued)

Table 3. (Continued)

TERM	P	CHEMICAL FORMULA	GENES	STRUCTURAL COMPOSITION
Progesterone CTD 00006624	3.44E-06	C <sub>21</sub> H <sub>30</sub> O <sub>2</sub>	<i>PTGFR</i> ; <i>SLC26A2</i> ; <i>CSTA</i> ; <i>CRABP2</i> ; <i>CFH</i> , etc	
Cyclophosphamide CTD 00005734	1.84E-05	C <sub>7</sub> H <sub>15</sub> Cl <sub>2</sub> N <sub>2</sub> O <sub>2</sub> P	<i>NKAIN1</i> ; <i>TMEFF1</i> ; <i>CSTA</i> ; <i>CRABP2</i> ; <i>CYP2D6</i> , etc	
Sertraline CTD 00007358	3.00E-05	C <sub>17</sub> H <sub>17</sub> Cl <sub>2</sub> N	<i>FABP1</i> ; <i>SLC5A7</i> ; <i>CYP2D6</i> ; <i>DUSP1</i> ; <i>CASP4</i> ; <i>HSPA1B</i> ; <i>HSPA1A</i>	
Phenytoin CTD 00006527	3.00E-05	C <sub>15</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub>	<i>EOMES</i> ; <i>COL1A1</i> ; <i>COL1A2</i> ; <i>CYP2D6</i> ; <i>COL6A2</i> ; <i>COL6A3</i> ; <i>S100A4</i>	
Menadione PC3 UP	3.78E-05	C <sub>11</sub> H <sub>8</sub> O <sub>2</sub>	<i>OSR2</i> ; <i>DUSP1</i> ; <i>CYP1B1</i> ; <i>MSX1</i> ; <i>KLF4</i> ; <i>HSPA1B</i> ; <i>HSPA1A</i>	

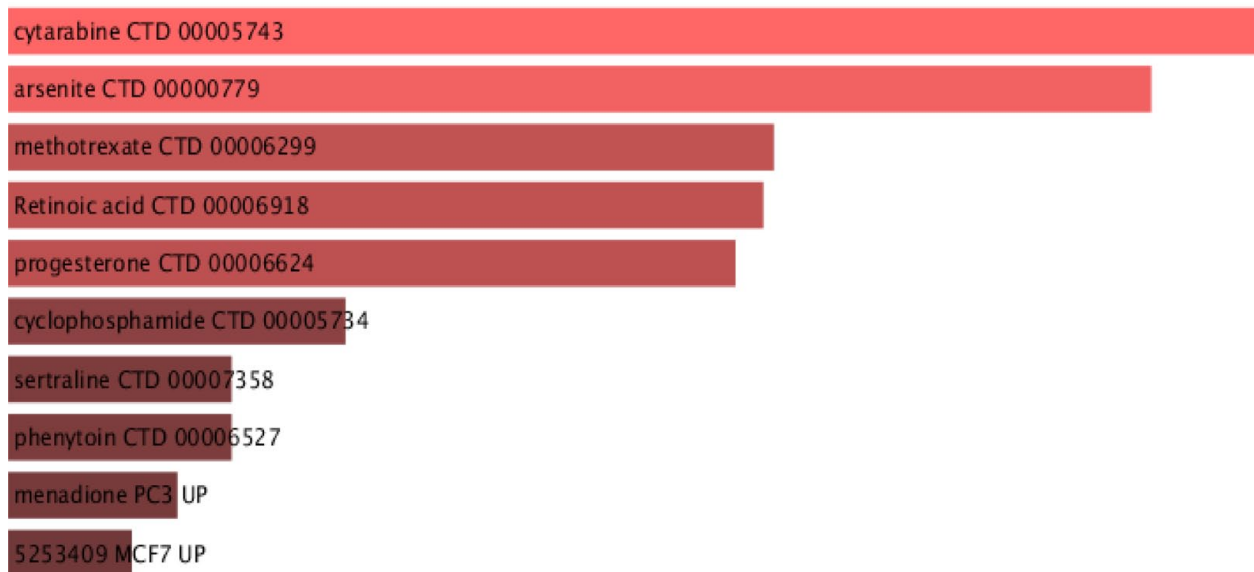
as a hub protein.<sup>73,74</sup> S100A4 has been shown in 2 previous studies to be a useful biomarker and to be abundantly expressed in HD.<sup>75,76</sup> We found MSX1 to be a potential hub protein which is relevant to some studies.<sup>77,78</sup>

## Discussion

Huntington disease is a hereditary neurodegenerative condition which impacts in early forties and fifties.<sup>24,79</sup> Dysregulation of possible biomarkers has recently been found to be linked with neurological and neurodegenerative diseases,<sup>80,81</sup> and several studies have looked into the implications of biomolecular activities in patients with HD.<sup>27,82-85</sup> Omics-based approaches are currently being used more often in biomedicine and systemic biology studies, and they have proven to be a helpful resource for deciphering disease pathogenesis, discovering molecular pathways, and determining biomarkers for various diseases.<sup>31</sup> Gene expression is perhaps modulated at multiple stages of RNA processing, post-translational modification (PTM) of proteins, translation, or other genomic changes, according to previous researches.<sup>86,87</sup> Determining the

biochemical pathway of an exact disease and identifying the impact of fundamental mechanisms involved in a certain phenotype require defining the activity target proteins in bioactive molecules.<sup>31,88</sup> The functional characteristics of hub proteins are of special intrigue, and PPI can be characterized as persistent or transient, depending on how long they last and what they do.<sup>89</sup> The networks based on PPI are considered scale-free where the connectivity between components follows generally a Poisson distribution.<sup>90</sup> Adapting a network-based technique to genomic data aids in the discovery of connections between diverse biological activities and processes, resulting in new pathways, interaction networks, as well as other disease-related signals that result in identifying biomarkers and treatment targets.<sup>91</sup> Although several studies identified miRNA expression of cellular and mouse structure,<sup>85,92</sup> common gene expression of HD patient,<sup>82</sup> and DNA methylation in HD,<sup>27</sup> a bioinformatics study regarding the identification of molecular signatures and pathways of healthy controls and HD patients has yet not been conducted comprehensively in a framework comparing all these methods at same time. We used a comprehensive





**Figure 7.** Bar graph including top 10 related small molecules with high significant correlations according to *P* values.

bioinformatics strategy to identify molecular signatures and key pathways for HD in this study and emphasized an overall overview.

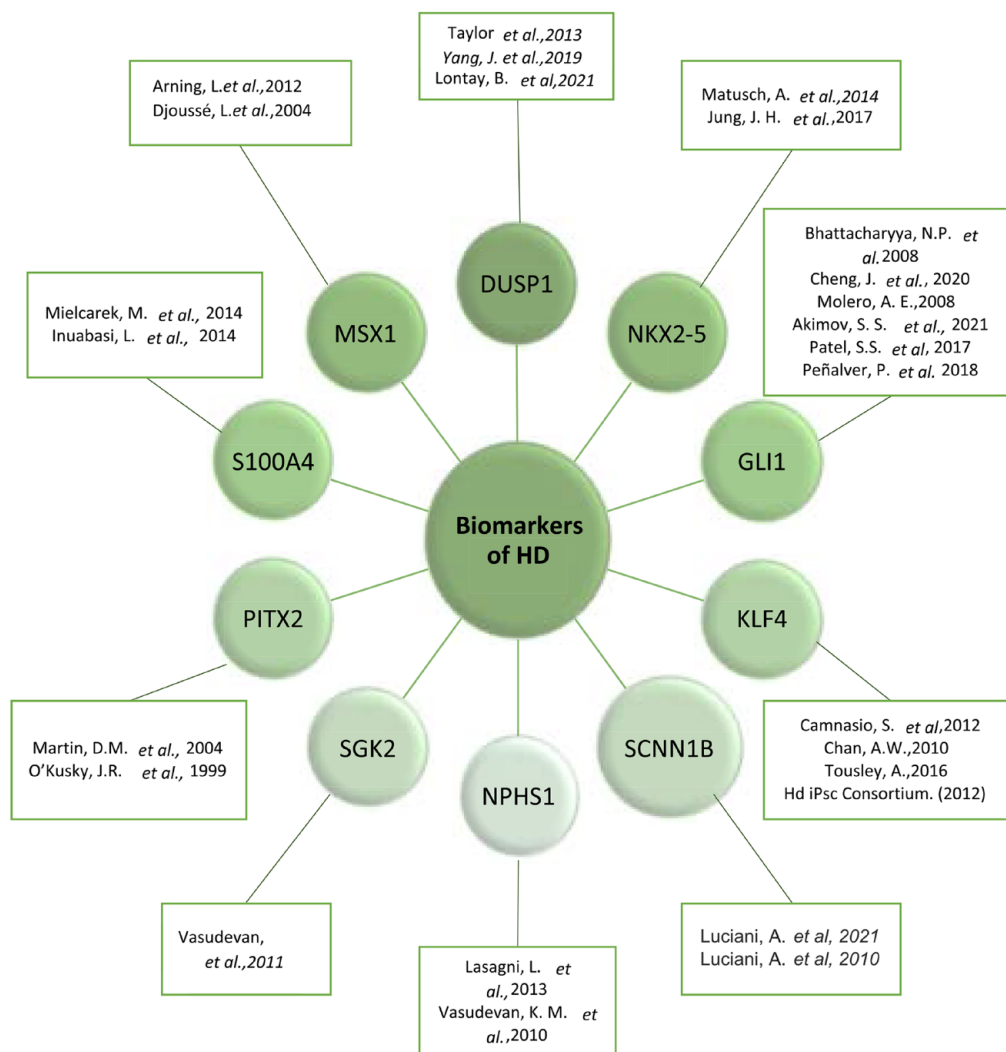
The key word ‘Huntington’s Disease’ was used to explore the GEO database, and data sets featuring *Homo sapiens* mRNA and array expression profiles were sorted containing control and diseased groups after reviewing some literatures.<sup>28</sup> A bioinformatics approach was used to assess DEGs between the control and diseased groups where the HD patients revealed significant variations in gene expression compared with neurologically healthy controls. The selected databases (GSE64810 and GSE95343) showed total 162 common genes among them 106 genes were found up-regulated and 56 genes were found down-regulated by systematic and statistical approaches. The data set GSE95343 has not been studied yet in the view of comprehensive bioinformatics, and it was found that GSE64810 was studied separately in a study where 1612 DEGs had been determined.<sup>83</sup>

The GO analysis was used to evaluate the biological relevance of HD regarding 162 common DEGs found. These important genes are identified to generate proteins with various molecular functions linked with important biological processes such as skeletal system development, collagen fibril organization, embryonic forelimb morphogenesis, proteolysis, and positive regulation of tumor necrosis factor-mediated signaling pathway matrix as well. However, GO term signal transduction in Biological Process<sup>83,84</sup> and plasma membrane in Cellular Component<sup>93</sup> were found in some other related researches. In our study, TF activity and sequence-specific DNA binding was found in molecular function enriched by gene ontologies, whereas TF activity in molecular function<sup>3</sup> was found to be relevant to HD in a study. Combining earlier study findings, the current analysis may give new treatment targets or probable pathogenesis to be investigated further.

In our study, the hub genes we identified were *DUSP1*, *NKX2-5*, *GLI1*, *KLF4*, *SCNN1B*, *NPHS1*, *SGK2*, *PITX2*, *S100A4*, and *MSX1* reflecting as the potential biomarkers for HD. These hub proteins are thought to be important particle in the disease-causing pathways.<sup>94</sup> As a result, we recreated the protein interaction network with a focus on DEGs to find relevant hub proteins. These proteins could play a role in the development and progression of HD. In HD, sodium butyrate has previously been shown to be neuroprotective, and this is correlated with an increased affirmation of mitogen-activated protein kinase phosphatase 1 (*DUSP1/MKP1*).<sup>53</sup> In some diseases, such as HD, low *DUSP1* (also known as *MKP-1*) expression may contribute to *MAPK* hyperactivation, whereas increased *MKP* activity may be neuroprotective.<sup>53-55</sup> *GLI1* has also been identified as a contributing hub protein gene in HD, which has been linked to a number of other studies.<sup>58-63</sup>

In a study, *NKX2-5* was discovered to be associated with protein synthesis in HD along with *AGG* and *GATA-4*.<sup>56</sup> *NKX2-5* was also found as one of the significantly ( $P < .05$ ) up-regulated genes.<sup>57</sup> Previous research has suggested *KLF4* as a possible biomarker for HD, which is supported by our findings.<sup>64-68</sup> Histones and nephrin (*NPHS1*) levels in the neuronal cytoplasmic pool were found to be higher in HD and Alzheimer disease brains in a previous study. Increased *NPHS1* expression has been linked to HD in a study, which backs up our findings.<sup>71</sup> *SCNN1B*s were identified as a molecular biomarker in HD which was also demonstrated in 2 earlier studies.<sup>69,70</sup>

Decreased *AKT* signaling has been linked to the pathophysiology of HD in a previous study, and the *SGK* family has gained a lot of interest because of its high similarity to *AKT*.<sup>72</sup> A study identified that in mammalian genomes, 3 genes encode the *SGK* family (*SGK1*, *SGK2*, and *SGK3*).<sup>72</sup> In our research, we discovered that *SGK2* is a possible protein hub in HD



**Figure 8.** Graphical representation of the validation of relevant biomarkers.

which is relevant to this previous study. Loss of PITX2 in HD does not affect the initial formation of striatal neurons composing the pathways directly and indirectly, according to recent studies although over-expressed PITX2 affects neurons.<sup>73,74</sup> In mice models of HD, *S100A4* genes were found to be considerably up-regulated within early-symptomatic R6/2 as well as pre-symptomatic HdhQ150 hearts, according to a study,<sup>76</sup> and it was also identified as up-regulated in another research.<sup>75</sup>

Directory biomolecules are being explored more and more as potential biomarkers for serious illnesses like neurodegenerative disorders.<sup>26,31,95,96</sup> Keeping this in consideration, we investigated the involvement of TFs and miRNAs in DEG regulation via TF-miRNA coregulatory networks. MicroRNAs have a crucial role in the regulation of gene expression, and there is growing evidence that they could be used as biomarkers for HD and other disorders.<sup>83,92</sup> Many miRNAs are expected to have a role in the pathogenic problem that introduces HD.<sup>2,94</sup> Our study disclosed top significant transcriptomic factors (TFs), ie, FOXC1, GATA2, PRRX2, YY1, TFAP2A, FOXL1, PPARG, HINFP, STAT3, MEF2A, and NFKB1 and

top significant miRNAs, ie, hsa-miR-340, hsa-miR-34a, hsa-miR-495, hsa-miR-1, hsa-miR-124, hsa-miR-29a, hsa-miR-29b, hsa-miR-30e, hsa-miR-16, hsa-miR-206, hsa-miR-30a, hsa-miR-30c, and hsa-miR-944 as regulatory biomarkers for HD. FOXC1, GATA2, YY1, FOXL1, and PPARG, which were revealed to be engaged as regulatory TFs in another neurodegenerative disease (Alzheimer disease, Parkinson disease, and some other neurodegenerative diseases also), were among the TFs we detected.<sup>81,95,96</sup> In a study, researchers discovered an increase in the concentration of TFAP2A nucleoid signals from various micropattern colonies in HD.<sup>97</sup> Some relevant studies showed similar result as our findings resulting the TFs PPARG, STAT3, MEF2A, and NFKB1 be found in HD side by side other TFs.<sup>22,98-100</sup>

It has been found that hsa-miR-340, hsa-miR-34a, hsa-miR-495, hsa-miR-30e, hsa-miR-206, and hsa-miR-30a have contribution to as a down-regulated miRNA;<sup>96,101-104</sup> on the contrary, hsa-miR-1 was found to have potentiality in Schizophrenia.<sup>105</sup> Huntington disease, Alzheimer disease, Hypoxic Ischemic Encephalopathy,

Parkinson disease, and ischemic stroke have all been identified to have aberrant miR-124 expression.<sup>106,107</sup> Our findings revealed that decreased expression of miR-29a and hsa-miR-29b is a prevalent occurrence in many neurodegenerative disorders, including HD<sup>85</sup> and Alzheimer disease<sup>108</sup> and hsa-miR-16 is also a potential miRNA for HD.<sup>109</sup> Differentially expressed genes were then investigated further to understand more about the prospective interactions using minor candidate drugs that could cure HD.

Finally, we predicted drugs/compounds to produce drug repositioning hypothesis in HD based on the probability that revealed biomarkers (ie, hub proteins and TFs) could be potential therapeutic targets.<sup>110–112</sup> As a result, connections between identified HD biomarkers and medicines have been established, suggesting that they may impact critical pathways in disease development<sup>113</sup>; nevertheless, more research is needed to assess the ramifications of proposed biomarker blocking.

## Conclusions

The successful implementation of disease-modifying therapies in HD progression will require the detection of potential biomarkers and their pathways. This research provided a summary of network-based approaches for identifying biochemical mechanisms underlying HD progression. From 2 transcriptomic data sets of HD, 162 DEGs were identified using a comprehensive bioinformatics analysis to build a PPI network, and the top most significant hub genes from the PPI network were identified as potential novel biomarkers in HD diagnosis. Following that, several TFs, ie, FOXC1, GATA2, PRRX2, etc and miRNAs such as, hsa-miR-340, hsa-miR-34a, hsa-miR-495, hsa-miR-1, etc were identified from the common genes linked to HD. However, more research is needed to confirm the anticipated drugs. We anticipate that all these biomarkers will allow quicker and more affordable detection of brain samples for the identification of HD. This method of identifying biomarkers can be used in obtainable tissue from brain to assess their presence and activity in inaccessible tissue, and it could be used to solve other therapeutic difficulties. We therefore suggest a more thorough validation of this framework and the possible biomarker transcripts we detected through clinical studies.

## Author contributions

MHR and MAA conceived and designed the study; TMM, UK and MSK performed the computational analyses; TMM and UK wrote the draft manuscript; UK and MBRM helped in the preparation of the tables and figures; MAA, MSK and MHR were involved in preparing the important intellectual content and critical revision; MAA, MHR and MSK supervised the whole study. All authors approved the final version for submission.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

All authors gave their consent for publication.

## Availability of data and materials

Not applicable.

## REFERENCES

- McColgan P, Tabrizi SJ. Huntington's disease: a clinical review. *Eur J Neurol*. 2018;25:24–34. doi:10.1111/ene.13413
- Moss DJH, Pardiñas AF, Langbehn D, et al. Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study [published correction appears in *Lancet Neurol*. 2017;16(9):683]. *Lancet Neurol*. 2017;16:701–711. doi:10.1016/S1474-4422(17)30161-8
- van Hagen M, Piebes DGE, de Leeuw WC, et al. The dynamics of early-state transcriptional changes and aggregate formation in a Huntington's disease cell model. *BMC Genomics*. 2017;18:373. doi:10.1186/s12864-017-3745-z
- Xiang C, Cong S, Liang B, Cong S. Bioinformatic gene analysis for potential therapeutic targets of Huntington's disease in pre-symptomatic and symptomatic stage. *J Transl Med*. 2020;18:388. doi:10.1186/s12967-020-02549-9
- Khan H, Ullah H, Tundis R, et al. Dietary flavonoids in the management of Huntington's disease: mechanism and clinical perspective. *eFood*. 2020;13:38–52. doi:10.2991/efood.k.200203.001
- Ghosh R, Tabrizi SJ. Clinical features of Huntington's disease. In: *Polyglutamine Disorders*. 2018:1–28. doi:10.1007/978-3-319-71779-1\_1
- García TP, Marder K. Statistical approaches to longitudinal data analysis in neurodegenerative diseases: Huntington's disease as a model. *Curr Neurol Neurosci Rep*. 2017;17:14. doi:10.1007/s11910-017-0723-4
- Myers RH, Sax DS, Schoenfeld M, et al. Late onset of Huntington's disease. *J Neurol Neurosurg Psychiatry*. 1985;48:530–534. doi:10.1136/jnnp.48.6.530
- Shannon KM. Huntington's disease—clinical signs, symptoms, presymptomatic diagnosis, and diagnosis. *Handb Clin Neurol*. 2011;100:3–13. doi:10.1016/B978-0-444-52014-2.00001-X
- Carbo M, Brandi V, Pascarella G, et al. Bioinformatics analysis of Ras homologue enriched in the striatum, a potential target for Huntington's disease therapy. *Int J Mol Med*. 2019;44:2223–2233. doi:10.3892/ijmm.2019.4373
- Azambuja MJ, Haddad MS, Radanovic M, Barbosa ER, Mansur LL. Semantic, phonologic, and verb fluency in Huntington's disease. *Dement Neuropsychol*. 2007;1:381–385. doi:10.1590/S1980-57642008DN10400009
- Novak MJ, Tabrizi SJ. Huntington's disease: clinical presentation and treatment. *Int Rev Neurobiol*. 2011;98:297–323. doi:10.1016/B978-0-12-381328-2.00013-4
- Davenport CB. Huntington's chorea in relation to heredity and eugenics. *Proc Natl Acad Sci U S A*. 1915;1:283–285. doi:10.1073/pnas.1.5.283
- Harper PS. The epidemiology of Huntington's disease. *Hum Genet*. 1992;89:365–376. doi:10.1007/BF00194305
- Rawlins MD, Wexler NS, Wexler AR, et al. The prevalence of Huntington's disease. *Neuroepidemiology*. 2016;46:144–153. doi:10.1159/000443738
- Scrimgeour EM, Pfuojena JW. Huntington disease in black Zimbabwean families living near the Mozambique border. *Am J Med Genet*. 1992;44:762–766. doi:10.1002/ajmg.1320440610
- Baig SS, Strong M, Quarrell OW. The global prevalence of Huntington's disease: a systematic review and discussion. *Neurodegener Dis Manag*. 2016;6:331–343. doi:10.2217/nmt-2016-0008
- Labbadia J, Morimoto RI. Huntington's disease: underlying molecular mechanisms and emerging concepts. *Trends Biochem Sci*. 2013;38:378–385. doi:10.1016/j.tibs.2013.05.003
- McFarland KN, Cha JH. Molecular biology of Huntington's disease. *Handb Clin Neurol*. 2011;100:25–81. doi:10.1016/B978-0-444-52014-2.00003-3
- Dickey AS, La Spada AR. Therapy development in Huntington disease: from current strategies to emerging opportunities. *Am J Med Genet A*. 2018;176:842–861. doi:10.1002/ajmg.a.38494
- Bondi MW, Edmonds EC, Salmon DP. Alzheimer's disease: past, present, and future. *J Int Neuropsychol Soc*. 2017;23:818–831. doi:10.1017/S135561771700100X
- Dong X, Cong S. Identification of differentially expressed genes and regulatory relationships in Huntington's disease by bioinformatics analysis. *Mol Med Rep*. 2018;17:4317–4326. doi:10.3892/mmr.2018.8410
- Zadel M, Maver A, Kovanda A, Peterlin B. DNA methylation profiles in whole blood of Huntington's disease patients. *Front Neurol*. 2018;9:655. doi:10.3389/fneur.2018.00655



24. Wang ZM, Dong XY, Cong SY. Bioinformatic analysis of a microRNA regulatory network in Huntington's disease. *J Integr Neurosci*. 2020;19:641-650. doi:10.31083/j.jin.2020.04.203
25. Paulsen JS. Cognitive impairment in Huntington disease: diagnosis and treatment. *Curr Neurol Neurosci Rep*. 2011;11:474-483. doi:10.1007/s11910-011-0215-x
26. Mastrokolias A, Ariyurek Y, Goeman JJ, et al. Huntington's disease biomarker progression profile identified by transcriptome sequencing in peripheral blood. *Eur J Hum Genet*. 2015;23:1349-1356. doi:10.1038/ejhg.2014.281
27. Thomas EA. DNA methylation in Huntington's disease: implications for trans-generational effects. *Neurosci Lett*. 2016;625:34-39. doi:10.1016/j.neulet.2015.10.060
28. Home-GEO-NCBI [Internet]. Accessed November 7, 2021. <https://www.ncbi.nlm.nih.gov/geo/>
29. Chowdhury UN, Ahmad S, Islam MB, et al. System biology and bioinformatics pipeline to identify comorbidities risk association: neurodegenerative disorder case study. *PLoS ONE*. 2021;16:e0250660. doi:10.1371/journal.pone.0250660
30. Bristy SA, Islam AH, Andalib KS, Khan U, Awal MA, Rahman MH. Determination of molecular signatures and pathways common to brain tissues of autism spectrum disorder: insights from comprehensive bioinformatics approach. *Inform Med Unlocked*. 2022;29:100871. doi:10.1016/j.imu.2022.100871
31. Rahman MR, Islam T, Huq F, Quinn JM, Moni MA. Identification of molecular signatures and pathways common to blood cells and brain tissue of amyotrophic lateral sclerosis patients. *Inform Med Unlocked*. 2019;16:100193. doi:10.1016/j.imu.2019.100193
32. Phipson B, Lee S, Majewski JJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann Appl Stat*. 2016;10:946-963. doi:10.1214/16-AOAS920
33. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57:289-300. doi:10.1111/j.2517-6161.1995.tb02031.x
34. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545-15550. doi:10.1073/pnas.0506580102
35. Khan U, Habibur Rahman M, Salauddin Khan M, Hossain MS, Morsaline Billah M. Bioinformatics and network-based approaches for determining pathways, signature molecules, and drug substances connected to genetic basis of schizophrenia etiology. *Brain Res*. 2022;1785:147889. doi:10.1016/j.brainres.2022.147889
36. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90-W97. doi:10.1093/nar/gkw377
37. Mahi NA, Najafabadi MF, Pilarczyk M, Kouril M, Medvedovic M. GREIN: an interactive web platform for re-analyzing GEO RNA-seq data. *Sci Rep*. 2019;9:7580. doi:10.1038/s41598-019-43935-8
38. David F. Waugh, protein-protein interactions. In: Anson, ML, Bailey, K, Edsall, JT, eds. *Advances in Protein Chemistry*. Academic Press; 1954:325-437. doi:10.1016/S0065-3233(08)60210-7
39. Nooren IM, Thornton JM. Diversity of protein-protein interactions. *EMBO J*. 2003;22:3486-3492. doi:10.1093/emboj/cdg359
40. Xia J, Gill EE, Hancock RE. Network analyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc*. 2015;10:823-844. doi:10.1038/nprot.2015.052
41. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011;27:431-432. doi:10.1093/bioinformatics/btq675
42. Papavassiliou AG. Transcription factors. *New Engl J Med*. 1995;332:45-47. doi:10.1056/NEJM199501053320108
43. Islam AMH, Rahman MH, Bristy SA, et al. Identification of molecular signatures and pathways common to blood cells and brain tissue-based RNA-Seq datasets of bipolar disorder: insights from comprehensive bioinformatics approach. *Inform Med Unlocked*. 2022;29:100881. doi:10.1016/j.imu.2019.100193
44. JASPAR: an open-access database of transcription factor binding profiles [Internet]. Accessed April 6, 2022. <http://jaspar.genereg.net>
45. TarBase v8 [Internet]. Accessed April 6, 2022. <https://bio.tools/tarbase>. <https://bio.tools/tarbase>
46. Goldman SM. Environmental toxins and Parkinson's disease. *Annu Rev Pharmacol Toxicol*. 2014;54:141-164. doi:10.1146/annurev-pharmtox-011613-135937
47. Zhou H, Skolnick J. A knowledge-based approach for predicting gene-disease associations. *Bioinformatics*. 2016;32:2831-2838. doi:10.1093/bioinformatics/btw358
48. DisGeNET—a database of gene-disease associations [Internet]. Accessed April 6, 2022. <https://www.disgenet.org/>
49. DSigDB—Tan Lab [Internet]. Accessed April 6, 2022. <http://dsigdb.tanlab.org/DSigDBv1.0/>
50. PubChem. Accessed May 16, 2022. <https://pubchem.ncbi.nlm.nih.gov/>
51. Google scholar [Internet]. Accessed November 29, 2021. <https://scholar.google.com/>
52. PubMed [Internet]. Accessed November 29, 2021. <https://pubmed.ncbi.nlm.nih.gov/>
53. Taylor DM, Moser R, Régulier E, et al. MAP kinase phosphatase 1 (MKP-1/DUSP1) is neuroprotective in Huntington's disease via additive effects of JNK and p38 inhibition. *J Neurosci*. 2013;33:2313-2325. doi:10.1523/JNEUROSCI.4965-11.2013
54. DUSP1/MKP-1 regulates proliferation and apoptosis in keratinocytes through the ERK/Elk-1/Egr-1 signaling pathway. *Life Sci*. 2019;223:47-53.
55. Lontay B, Kiss A, Virág L, Tar K. How do post-translational modifications influence the pathomechanistic landscape of Huntington's disease? a comprehensive review. *Int J Mol Sci*. 2020;21:4282. doi:10.3390/ijms21124282
56. Matusch A, Saft C, Elmenhorst D, et al. Cross sectional PET study of cerebral adenosine A<sub>1</sub> receptors in premanifest and manifest Huntington's disease. *Eur J Nucl Med Mol Imaging*. 2014;41:1210-1220. doi:10.1007/s00259-014-2724-8
57. Jung JH, Ko J, Lee EH, et al. RNA seq- and DEG-based comparison of developmental toxicity in fish embryos of two species exposed to Iranian heavy crude oil. *Comp Biochem Physiol C Toxicol Pharmacol*. 2017;196:1-10. doi:10.1016/j.cbpc.2017.02.010
58. Bhattacharyya NP, Banerjee M, Majumder P. Huntington's disease: roles of Huntingtin-interacting protein 1 (HIP-1) and its molecular partner HIPPI in the regulation of apoptosis and transcription. *FEBS J*. 2008;275:4271-4279. doi:10.1111/j.1742-4658.2008.06563.x
59. Cheng J, Liu HP, Lin WY, Tsai FJ. Identification of contributing genes of Huntington's disease by machine learning. *BMC Med Genomics*. 2020;13:176. doi:10.1186/s12920-020-00822-w
60. Molero AE. Neurodevelopmental abnormalities in an animal model of Huntington's disease. Published 2008. Accessed November 3, 2021. <https://repository.yu.edu/handle/20.500.12202/934>
61. Akimov SS, Jiang M, Kedaigle AJ, et al. Immortalized striatal precursor neurons from Huntington's disease patient-derived iPSCs as a platform for target identification and screening for experimental therapeutics. *Hum Mol Genet*. 2021;30:2469-2487. doi:10.1093/hmg/ddab200
62. Patel SS, Tomar S, Sharma D, Mahindroo N, Udayabanu M. Targeting sonic hedgehog signaling in neurological disorders. *Neurosci Biobehav Rev*. 2017;74:76-97. doi:10.1016/j.neubiorev.2017.01.008
63. Peñalver P, Belmonte-Reche E, Adán N, et al. Alkylated resveratrol prodrugs and metabolites as potential therapeutics for neurodegenerative diseases. *Eur J Med Chem*. 2018;146:123-138. doi:10.1016/j.ejmech.2018.01.037
64. Camnasio S, Delli Carri A, Lombardo A, et al. The first reported generation of several induced pluripotent stem cell lines from homozygous and heterozygous Huntington's disease patients demonstrates mutation related enhanced lysosomal activity. *Neurobiol Dis*. 2012;46:41-51. doi:10.1016/j.nbd.2011.12.042
65. Chan AW, Cheng PH, Neumann A, Yang JJ. Reprogramming Huntington monkey skin cells into pluripotent stem cells. *Cell Reprogram*. 2010;12:509-517. doi:10.1089/cell.2010.0019
66. Tousley A, Kegel-Gleason KB. Induced pluripotent stem cells in Huntington's disease research: progress and opportunity. *J Huntingtons Dis*. 2016;5:99-131. doi:10.3233/JHD-160199
67. HD iPSC Consortium. Induced pluripotent stem cells from patients with Huntington's disease show CAG-repeat-expansion-associated phenotypes. *Cell Stem Cell*. 2012;11(2):264-278. doi:10.1016/j.stem.2012.04.027
68. Jeon I, Lee N, Li JY, et al. Neuronal properties, in vivo effects, and pathology of a Huntington's disease patient-derived induced pluripotent stem cells [published correction appears in *Stem Cells*. 2012;30(11):2602]. *Stem Cells*. 2012;30:2054-2062. doi:10.1002/stem.1135
69. Luciani A, Vilella VR, Esposito S, et al. Cystic fibrosis: a disorder with defective autophagy. *Autophagy*. 2011;7:104-106. doi:10.4161/auto.7.1.13987
70. Luciani A, Vilella VR, Esposito S, et al. Defective CFTR induces aggregate formation and lung inflammation in cystic fibrosis through ROS-mediated autophagy inhibition. *Nat Cell Biol*. 2010;12:863-875. doi:10.1038/ncb2090
71. Lasagni L, Lazzeri E, Shankland SJ, Anders HJ, Romagnani P. Podocyte mitosis—a catastrophe. *Curr Mol Med*. 2013;13:13-23. doi:10.2174/1566524011307010013
72. Vasudevan KM, Garraway LA. AKT signaling in physiology and disease. *Curr Top Microbiol Immunol*. 2010;347:105-133. doi:10.1007/82\_2010\_66
73. Martin DM, Skidmore JM, Philips ST, et al. PITX2 is required for normal development of neurons in the mouse subthalamic nucleus and midbrain. *Dev Biol*. 2004;267:93-108. doi:10.1016/j.ydbio.2003.10.035
74. O'Kusky JR, Nasir J, Cicchetti F, Parent A, Hayden MR. Neuronal degeneration in the basal ganglia and loss of pallido-subthalamic synapses in mice with targeted disruption of the Huntington's disease gene. *Brain Res*. 1999;818:468-479. doi:10.1016/s0006-8993(98)01312-2
75. Mielcarek M, Bondulich MK, Inuabasi L, Franklin SA, Muller T, Bates GP. The Huntington's disease-related cardiomyopathy prevents a hypertrophic response in the R6/2 mouse model. *PLoS ONE*. 2014;9:e108961. doi:10.1371/journal.pone.0108961



76. Mielcarek M, Inuabasi L, Bondulich MK, et al. Dysfunction of the CNS-heart axis in mouse models of Huntington's disease. *PLoS Genet.* 2014;10:e1004550. doi:10.1371/journal.pgen.1004550
77. Arning L, Epplen JT. Genetic modifiers of Huntington's disease: beyond CAG. *Future Neurol.* 2012;7:93-109.
78. Djoussé L, Knowlton B, Hayden MR, et al. Evidence for a modifier of onset age in Huntington disease linked to the HD gene in 4p16. *Neurogenetics.* 2004;5:109-114.
79. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498-2504. doi:10.1101/gr.1239303
80. Junn E, Mouradian MM. MicroRNAs in neurodegenerative diseases and their therapeutic potential. *Pharmacol Ther.* 2012;133:142-150. doi:10.1016/j.pharmthera.2011.10.002
81. Molasy M, Walczak A, Szaflik J, Szaflik JP, Majsterek I. MicroRNAs in glaucoma and neurodegenerative diseases. *J Hum Genet.* 2017;62:105-112. doi:10.1038/jhg.2016.91
82. Neueder A, Bates GP. A common gene expression signature in Huntington's disease patient brain regions. *BMC Med Genomics.* 2014;7:60. doi:10.1186/s12920-014-0060-2
83. Dong X, Cong S. Bioinformatic analysis of microRNA expression in Huntington's disease. *Mol Med Rep.* 2018;18:2857-2865. doi:10.3892/mmr.2018.9238
84. Träger U, Andre R, Lahiri N, et al. HTT-lowering reverses Huntington's disease immune dysfunction caused by NFκB pathway dysregulation. *Brain.* 2014;137:819-833. doi:10.1093/brain/awt355
85. Johnson R, Zuccato C, Belyaev ND, Guest DJ, Cattaneo E, Buckley NJ. A microRNA-based gene dysregulation pathway in Huntington's disease. *Neurobiol Dis.* 2008;29:438-445. doi:10.1016/j.nbd.2007.11.001
86. Cha JH. Transcriptional signatures in Huntington's disease. *Prog Neurobiol.* 2007;83:228-248. doi:10.1016/j.pneurobio.2007.03.004
87. Hwang YJ, Hyeon SJ, Kim Y, et al. Modulation of SETDB1 activity by APQ ameliorates heterochromatin condensation, motor function, and neuropathology in a Huntington's disease mouse model. *J Enzyme Inhib Med Chem.* 2021;36:856-868. doi:10.1080/14756366.2021.1900160
88. Hwang HY, Kim TY, Szász MA, et al. Profiling the protein targets of unmodified bio-active molecules with drug affinity responsive target stability and liquid chromatography/tandem mass spectrometry. *Proteomics.* 2020;20:e1900325. doi:10.1002/pmic.201900325
89. Reynolds C, Damerell D, Jones S. ProtorP: a protein-protein interaction analysis server. *Bioinformatics.* 2009;25:413-414. doi:10.1093/bioinformatics/btn584
90. Batada NN, Hurst LD, Tyers M. Evolutionary and physiological importance of hub proteins. *PLoS Comput Biol.* 2006;2:e88. doi:10.1371/journal.pcbi.0020088
91. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518:317-330. doi:10.1038/nature14248
92. Jin J, Cheng Y, Zhang Y, et al. Interrogation of brain miRNA and mRNA expression profiles reveals a molecular regulatory network that is perturbed by mutant huntingtin. *J Neurochem.* 2012;123:477-490. doi:10.1111/j.1471-4159.2012.07925.x
93. Shrivastava AN, Aperia A, Melki R, Triller A. Physico-pathologic mechanisms involved in neurodegeneration: misfolded protein-plasma membrane interactions. *Neuron.* 2017;95:33-50. doi:10.1016/j.neuron.2017.05.026
94. Sevimoglu T, Arga KY. The role of protein interaction networks in systems biomedicine. *Comput Struct Biotechnol J.* 2014;11:22-27. doi:10.1016/j.csbj.2014.08.008
95. Rahman MR, Islam T, Turanli B, et al. Network-based approach to identify molecular signatures and therapeutic agents in Alzheimer's disease. *Comput Biol Chem.* 2019;78:431-439. doi:10.1016/j.compbiolchem.2018.12.011
96. Rahman MR, Islam T, Zaman T, et al. Identification of molecular signatures and pathways to identify novel therapeutic targets in Alzheimer's disease: insights from a systems biomedicine perspective. *Genomics.* 2020;112:1290-1299. doi:10.1016/j.ygeno.2019.07.018
97. Piccolo FM, Kastan NR, Haremakei T, et al. Role of YAP in early ectodermal specification and a Huntington's disease model of human neurulation. *eLife.* 2022;11:e73075. doi:10.7554/eLife.73075
98. Ghose J, Sinha M, Das E, Jana NR, Bhattacharyya NP. Regulation of miR-146a by RelA/NFκB and p53 in STHdh(Q111)/Hdh(Q111) cells, a cell model of Huntington's disease. *PLoS ONE.* 2011;6:e23837. doi:10.1371/journal.pone.0023837
99. Mielcarek M, Mazur-Michalek I, Marcinkowski J, Wieleba M, Klimberg A, Zielonka D. Expression levels of MEF transcription factors in skeletal muscles of Huntington's disease mouse model [Poziom ekspresji czynników transkrypcyjnych MEF w mięśniach szkieletowych mysiego modelu choroby Huntingtona]. *Hygeia Pub Health.* 2019;54:64-67.
100. Chiang MC, Chen CM, Lee MR, et al. Modulation of energy deficiency in Huntington's disease via activation of the peroxisome proliferator-activated receptor gamma. *Hum Mol Genet.* 2010;19:4043-4058. doi:10.1093/hmg/ddq322
101. Li J, Xu C, Zhang J, et al. Identification of miRNA-target gene pairs in the parietal and frontal lobes of the brain in patients with Alzheimer's disease using bioinformatic analyses. *Neurochem Res.* 2021;46:964-979. doi:10.1007/s11064-020-03215-8
102. Su L, Chen S, Zheng C, Wei H, Song X. Meta-analysis of gene expression and identification of biological regulatory mechanisms in Alzheimer's disease. *Front Neurosci.* 2019;13:633. doi:10.3389/fnins.2019.00633
103. Lee ST, Chu K, Jung KH, et al. MmiR-206 regulates brain-derived neurotrophic factor in Alzheimer disease model. *Ann Neurol.* 2012;72:269-277. doi:10.1002/ana.23588
104. Lin F, Zhang H, Bao J, Li L. Identification of potential diagnostic miRNAs biomarkers for Alzheimer disease based on weighted gene coexpression network analysis. *World Neurosurg.* 2021;153:e315-e328. doi:10.1016/j.wneu.2021.06.118
105. Ardekani AM, Naeini MM. The role of MicroRNAs in human diseases. *Avicenna J Med Biotechnol.* 2010;2:161-179.
106. Ghafouri-Fard S, Shoori H, Bahroudi Z, Abak A, Majidpoor J, Taheri M. An update on the role of miR-124 in the pathogenesis of human disorders. *Biomed Pharmacother.* 2021;135:111198. doi:10.1016/j.biopha.2020.111198
107. Khan U, Khan MS. Prognostic value estimation of BRIP1 in breast cancer by exploiting transcriptomics data through bioinformatics approaches. *Bioinform Biol Insights.* 2021;15:11779322211055892. doi:10.1177/11779322211055892
108. Hébert SS, Horré K, Nicolai L, et al. Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. *Proc Natl Acad Sci U S A.* 2008;105:6415-6420. doi:10.1073/pnas.0710263105
109. Gauthwin PM, Ciesla M, Lahiri N, Tabrizi SJ, Brundin P, Björkqvist M. Hsa-miR-34b is a plasma-stable microRNA that is elevated in pre-manifest Huntington's disease. *Hum Mol Genet.* 2011;20:2225-2237. doi:10.1093/hmg/ddr111
110. Cho IK, Easley IVCA, Chan AW. Suppression of trinucleotide repeat expansion in spermatogenic cells in Huntington's disease. *J Assist Reprod Genet.* 2022;39:2413-2430.
111. Niewiadomska-Cimicka A, Krzyżosiak A, Ye T, et al. Genome-wide analysis of RARβ transcriptional targets in mouse striatum links retinoic acid signaling with Huntington's disease and other neurodegenerative disorders. *Mol Neurobiol.* 2017;54:3859-3878.
112. Duan W, Peng Q, Masuda N, et al. Sertraline slows disease progression and increases neurogenesis in N171-82Q mouse model of Huntington's disease. *Neurobiol Dis.* 2008;30:312-322.
113. Hernández-Espinosa D, Morton AJ. Calcineurin inhibitors cause an acceleration of the neurological phenotype in a mouse transgenic for the human Huntington's disease mutation. *Brain Res Bull.* 2006;69:669-679.