

# Gene-set enrichment analysis of differentially expressed genes of amniotic-fluid RNA providing insights into the phenotype of Turner's syndrome

Sarbojoy Saha<sup>1</sup>, Shampa Barmon<sup>2</sup>

1. Department of Biochemistry and Microbiology, North South University, Dhaka, Bangladesh.

\* Corresponding author. Email: [sarbojoy.saha@northsouth.edu](mailto:sarbojoy.saha@northsouth.edu)

2. Department of Environmental Science and Management, North South University, Dhaka, Bangladesh. Email: [shampab50@gmail.com](mailto:shampab50@gmail.com)

## Abstract

Genetic disorders are quite a major topic of discussion and debate in the recent world of biological sciences. Turner's syndrome is one such disorder caused by a chromosome aneuploidy and it has characteristic symptoms in the patient or the affected individual. The amniotic fluid is a complex biological material found in the amniotic sac of pregnant women and they can provide valuable knowledge and understanding of the pathogenesis of this particular chromosomal abnormality. In this study, global gene expression analysis of cell-free RNA in amniotic fluid supernatant was used to detect genes/organ systems which may be significant in the pathophysiology of Turner's syndrome. The cell-free RNA from the amniotic fluid of five mid-trimester Turner's syndrome fetuses and five euploid female fetuses matched for age of gestation were extracted, amplified and hybridized onto Affymetrix U33 Plus 2.0. array. The paired t-test was used to identify the significantly differentially regulated genes. Biological interpretation was conducted using ingenuity pathway analysis and BioGPS gene expression atlas. Of the genes, XIST was especially downregulated and SHOX was not expressed differentially. One of the most highly represented organ systems was the hematologic/immune system, differentiating the transcriptome of Turner's syndrome from other chromosomal aneuploidies that are discussed in this area of science. The differences in the transcriptome of the Turner's syndrome are due to genome-wide dysregulation. The hematologic/immune system differences are significant in early-onset autoimmune dysfunction. There are other genes which have been identified that are associated with the cardiovascular and the skeletal system, as these are often seen to be affected in the female patients with turner's syndrome. Hopefully, such knowledge gained from this study will help us to understand the deeper mechanisms of this disorder and the possible treatments of this disease.

**Keywords:** Turner's syndrome, chromosomal aneuploidy, amniotic fluid, cell-free RNA, Ingenuity pathway analysis

## Introduction

The subject of this study is the genetic disorder known as Turner's syndrome as described earlier in the abstract. It is a common sex chromosome aneuploidy with a relatively high prevalence (occurring 1 in 25000). Although in most of the cases (98-99 percent) the fetuses affected with Turner's syndrome end in miscarriages for the mother ("Fetology: Diagnosis and Management of the Fetal Patient : Mary E. D'Alton : 9780071442015" n.d.). Often, it is seen that women who has this disorder are normally unaware of it until they reach their adolescence and they learn about it much later even if they learn it somehow. This only makes it more difficult to put the spread of this disease in check. Turner's syndrome is due to monosomy X in 64 percent of prenatal cases and 47 percent of postnatal cases(Gravholt 2004). The remaining cases are because of other abnormalities such as deletions and duplications, ring chromosomes or mosaic aneuploidies. Turners syndrome often have characteristic signs such as short stature, webbed neck, coarctation of the aorta in the heart, lymphedema and infertility in the affected individual. Other symptoms include obesity, scoliosis, hyperlipidemia and juvenile rheumatoid arthritis(Tyler and Edman 2004). The abnormalities in the fetus include detectable by prenatal ultrasonography include increased nuchal translucency, hydrops and cystic hygroma(Papp et al. 2006).

The amniotic fluid found in the amniotic sac of pregnant women is routinely extracted and can provide a significant amount of understanding of the pathogenesis of chromosomal abnormalities of this type. It is often discarded but it is good source of cell-free RNA(Larrabee et al. 2005). It should be used for similar studies which could help us to understand the importance of this fluid as a subject material or a sample. It can give us knowledge about the effect of oxidative damage on genetic disorders(Slonim et al. 2009) (Koide et al. 2011). Early investigation of the common genes expressed in the amniotic fluid transcriptome has shown that several organ systems have been overrepresented such as the muscular-skeletal system and the hematologic or immune system(Hui et al. 2012).

In this particular study the global genes expression analysis was used to investigate the cell-free RNA transcriptome from amniotic fluid supernatants of fetuses affected by this genetic disorder, and this was done with the objective of understanding the pathological and physiological changes which are already present during the second trimester of the mother.

## Materials and methods

### Subjects/Ethical statement

The institutional review board at Tufts Medical Centre approved this research project. The unnamed/anonymous specimens were obtained from Integrated Genetics/LabCorp, located at Westborough, Maryland and is a leading research and diagnostic laboratory. The residual amniotic fluid samples (gestational age between 15 and 17/ 18 weeks) were collected from women undergoing routine genetic testing. The karyotype and gestational ages were the only information that were known, as the laboratory staff were blinded to ultrasound findings result.

The amniotic fluid was stored at a temperature of -80 degrees Celsius at Integrated Genetic, for a period of two to three months, transported with dry ice, and stored again at the same specific temperature at Tufts facility. Once the samples arrived at the testing facility, the RNA was extracted within a single week. The original sample consisted of seven 45, X samples which had a chromosome missing and seven 46, XX chromosomes which are normal having all the chromosomes. Due to poor quality of hybridization, two 45, X samples were left out from the analysis resulting in a cohort of five 45, X samples and five 46, XX samples. Sample sizes such as these have been seen to provide stability and reproducibility when used in other similar experiments. Hence, estimating the suitable sample size is very important when carrying out these types of studies as a lot depends on it.

### **Finding out the differentially expressed genes from GSE58435**

From the NCBI website GEO datasets were searched using the term “amniotic fluid RNA expression profiling for Turner’s syndrome and reference series GSE 58435 was analyzed with GEO2R.

For GEO2R analysis two groups named ‘Second trimester Turner’s syndrome fetus’ and ‘second trimester euploid fetus’ were defined. Five samples were included in each group. The GEOquery (Davis and Meltzer 2007) and limma R (Smyth n.d.) packages from the Bioconductor project were used, and GEO2R analysis was performed (Smyth 2004). Top 250 differentially expressed genes were found using the Benjamini & Hochberg (false discovery rate) method (Benjamini and Hochberg 1995) P values were adjusted. Log2 transformation to the data was applied. R script used to perform the calculation was obtained from the R script tab.

### **Enrichment Analysis**

Enrichment analysis of upregulated genes was carried out using

- ARCHS4 TFs co-expression (Lachmann et al. 2017)
- TRANSFAC and JASPAR PWMs (Khan et al. 2018)
- Reactome 2016 (Yu and He 2016)
- Biocarta 2016 (Chowdhury and Sarkar 2015)

Enrichment analysis of downregulated genes was carried out using

- ARCHS4 TFs co-expression (Lachmann et al. 2017)
- ARCHS4 Kinases Coexp (Lachmann et al. 2017)
- TRANSFAC and JASPAR PWMs (Khan et al. 2018)

## R Script

```
# Version info: R 3.2.3, Biobase 2.30.0, GEOquery 2.40.0, limma 3.26.8
# R scripts generated Wed Oct 10 05:24:57 EDT 2018

#####
# Differential expression analysis with limma
library(Biobase)
library(GEOquery)
library(limma)

# load series and platform data from GEO

gset <- getGEO("GSE58435", GSEMatrix =TRUE, AnnotGPL=TRUE)
if (length(gset) > 1) idx <- grep("GPL570", attr(gset, "names")) else idx <-
1
gset <- gset[[idx]]

# make proper column names to match toptable
fvarLabels(gset) <- make.names(fvarLabels(gset))

# group names for all samples
gsms <- "1111100000"
sml <- c()
for (i in 1:nchar(gsms)) { sml[i] <- substr(gsms,i,i) }

# log2 transform
ex <- exprs(gset)
qx <- as.numeric(quantile(ex, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))
LogC <- (qx[5] > 100) ||
        (qx[6]-qx[1] > 50 && qx[2] > 0) ||
        (qx[2] > 0 && qx[2] < 1 && qx[4] > 1 && qx[4] < 2)
if (LogC) { ex[which(ex <= 0)] <- NaN
  exprs(gset) <- log2(ex) }

# set up the data and proceed with analysis
sml <- paste("G", sml, sep="") # set group names
fl <- as.factor(sml)
gset$description <- fl
design <- model.matrix(~ description + 0, gset)
colnames(design) <- levels(fl)
fit <- lmFit(gset, design)
cont.matrix <- makeContrasts(G1-G0, levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2, 0.01)
tT <- topTable(fit2, adjust="fdr", sort.by="B", number=250)

tT <- subset(tT,
select=c("ID", "adj.P.Val", "P.Value", "t", "B", "logFC", "Gene.symbol", "Gene.title
"))
write.table(tT, file=stdout(), row.names=F, sep="\t")

#####
# Boxplot for selected GEO samples
library(Biobase)
```

## Results and Discussion

|  |                                      |          |                  |         |                |  |
|--|--------------------------------------|----------|------------------|---------|----------------|--|
| Upregulated genes                        |                                      |          |                  |         |                |  |
| Transcription                            |                                      |          |                  |         |                |  |
| <a href="#">ARCHS4 TFs Coexp</a>         |                                      |          |                  |         |                |  |
| Index                                    | Name                                 | P-value  | Adjusted p-value | Z-score | Combined score |  |
| 1  | ALX3_human_tf_ARCHS4_coexpression    | 0.002162 | 0.07256          | -2.26   | 13.84          |  |
| 2  | ZFP92_human_tf_ARCHS4_coexpression   | 0.07256  | 0.07256          | -2.25   | 5.91           |  |
| 3  | ZNF705A_human_tf_ARCHS4_coexpression | 0.07256  | 0.07256          | -2.1    | 5.52           |  |
| 4  | ZNF518A_human_tf_ARCHS4_coexpression | 0.07256  | 0.07256          | -1.97   | 5.16           |  |
| 5  | ZNF512_human_tf_ARCHS4_coexpression  | 0.07256  | 0.07256          | -1.91   | 5.01           |  |
| <a href="#">TRANSFAC and JASPAR PWMs</a> |                                      |          |                  |         |                |  |

| Index | Name             | P-value | Adjusted p-value | Z-score | Combined score |  |
|-------|------------------|---------|------------------|---------|----------------|--|
| 1     | E2F6 (human)     | 0.04014 | 0.4745           | -1.72   | 5.52           |  |
| 2     | ZBTB7A (human)   | 0.04096 | 0.4745           | -1.72   | 5.48           |  |
| 3     | TFAP2A (mouse)   | 0.05832 | 0.4745           | -1.66   | 4.72           |  |
| 4     | IRF8 (mouse)     | 0.06534 | 0.4745           | -1.63   | 4.44           |  |
| 5     | MZF1_1-4 (human) | 0.1383  | 0.4745           | -1.89   | 3.75           |  |

## Pathways

[Reactome 2016](#)

| Index | Name   | P-value  | Adjusted p-value | Z-score | Combined score |  |
|-------|--|----------|------------------|---------|----------------|--|
| 1     | Human Cytomegalovirus and Map Kinase Pathways_Homo sapiens_h_hcmvPathway                             | 0.003994 | 0.01022          | -1.41   | 7.8            |  |
| 2     | Phosphorylation of MEK1 by cdk5/p35 down regulates the MAP kinase pathway_Homo sapiens_h_cdk5Pathway | 0.003495 | 0.01022          | -1.35   | 7.66           |  |
| 3     | Cadmium induces DNA synthesis and proliferation in macrophages_Homo sapiens_h_cdMacPathway           | 0.003994 | 0.01022          | -1.38   | 7.6            |  |
| 4     | Melanocyte Development and Pigmentation Pathway_Homo sapiens_h_melanocytepathway                     | 0.003246 | 0.01022          | -1.15   | 6.62           |  |
| 5     | Regulation of Splicing through Sam68_Homo sapiens_h_sam68Pathway                                     | 0.002498 | 0.01022          | -1.04   | 6.21           |  |

[BioCarta 2016](#)

| Index | Name   | P-value  | Adjusted p-value | Z-score | Combined score |  |
|-------|--|----------|------------------|---------|----------------|--|
| 1     | Human Cytomegalovirus and Map Kinase Pathways_Homo sapiens_h_hcmvPathway | 0.003994 | 0.01022          | -1.41   | 7.8            |  |

|   |  |              |             |           |      |  |
|---|--|--------------|-------------|-----------|------|--|
| 2 | Phosphorylation of MEK1 by cdk5/p35 down regulates the MAP kinase pathway_Homo sapiens_h_cdk5Pathway | 0.0034<br>95 | 0.0102<br>2 | -<br>1.35 | 7.66 |  |
| 3 | Cadmium induces DNA synthesis and proliferation in macrophages_Homo sapiens_h_cdMacPathway           | 0.0039<br>94 | 0.0102<br>2 | -<br>1.38 | 7.6  |  |
| 4 | Melanocyte Development and Pigmentation Pathway_Homo sapiens_h_melanocytepathway                     | 0.0032<br>46 | 0.0102<br>2 | -<br>1.15 | 6.62 |  |
| 5 | Regulation of Splicing through Sam68_Homo sapiens_h_sam68Pathway                                     | 0.0024<br>98 | 0.0102<br>2 | -<br>1.04 | 6.21 |  |

| Downregulated genes                      |                                    |              |                  |           |                |
|--|------------------------------------|--------------|------------------|-----------|----------------|
| Transcription                            |                                    |              |                  |           |                |
| <a href="#">ARCHS4 TFs Coexp</a>         |                                    |              |                  |           |                |
| Index                                    | Name                               | P-value      | Adjusted p-value | Z-score   | Combined score |
| 1  | TCF21_human_tf_ARCHS4_coexpression | 0.0021<br>62 | 0.0131<br>3      | -<br>1.65 | 10.13          |
| 2  | OSR2_human_tf_ARCHS4_coexpression  | 0.0021<br>62 | 0.0131<br>3      | -<br>1.62 | 9.93           |
| 3  | OSR1_human_tf_ARCHS4_coexpression  | 0.0021<br>62 | 0.0131<br>3      | -<br>1.62 | 9.91           |
| 4  | TBX5_human_tf_ARCHS4_coexpression  | 0.0021<br>62 | 0.0131<br>3      | -<br>1.61 | 9.91           |
| 5  | PRRX1_human_tf_ARCHS4_coexpression | 0.0021<br>62 | 0.0131<br>3      | -<br>1.61 | 9.89           |
| 6  | ZHX3_human_tf_ARCHS4_coexpression  | 0.0021<br>62 | 0.0131<br>3      | -<br>1.61 | 9.86           |
| <a href="#">TRANSFAC and JASPAR PWMs</a> |                                    |              |                  |           |                |
| Index                                    | Name                               | P-value      | Adjusted p-value | Z-score   | Combined score |
| 1  | NR2C2 (human)                      | 0.0163<br>9  | 0.512            | -<br>2.18 | 8.96           |
| 2  | TBP (human)                        | 0.0157<br>9  | 0.512            | -1.6      | 6.66           |

| 3                                    | NFAT2 (human)                           | 0.04113  | 0.512            | -1.71   | 5.47           |
|--------------------------------------|---|----------|------------------|---------|----------------|
| 4                                    | PPARA (human)                           | 0.06595  | 0.512            | -1.93   | 5.26           |
| 5                                    | SRF (mouse)                             | 0.06614  | 0.512            | -1.6    | 4.35           |
| Pathways                             |   |          |                  |         |                |
| <a href="#">ARCHS4 Kinases Coexp</a> |   |          |                  |         |                |
| Index                                | Name                                    | P-value  | Adjusted p-value | Z-score | Combined score |
| 1                                    | HIPK3_human_kinase_ARCHS4_coexpression  | 0.002162 | 0.01135          | -1.71   | 10.52          |
| 2                                    | ALPK2_human_kinase_ARCHS4_coexpression  | 0.002162 | 0.01135          | -1.66   | 10.18          |
| 3                                    | DDR2_human_kinase_ARCHS4_coexpression   | 0.002162 | 0.01135          | -1.62   | 9.96           |
| 4                                    | PDGFRA_human_kinase_ARCHS4_coexpression | 0.002162 | 0.01135          | -1.61   | 9.88           |
| 5                                    | MYLK_human_kinase_ARCHS4_coexpression   | 0.002162 | 0.01135          | -1.61   | 9.86           |

It can be seen from the tabulated results that in the ARCHS4 TFs Coexpression database, ALX3\_human\_tf\_ARCHS4\_coexpression is the most significant transcription factor. And in the Reactome 2016 and Biocarta 2016 pathway, the Human Cytomegalovirus and Map Kinase Pathways\_Homo sapiens\_h\_hcmvPathway is the most significant pathway.

In the table showing the results for the downregulated pathway, the ARCHS4 TFs Coexpression database shows that TCF21\_human\_tf\_ARCHS4\_coexpression is the most significant transcription factor. And finally, in the ARCHS4 Kinases Coexpression database HIPK3\_human\_kinase\_ARCHS4\_coexpression is the most significant pathway.

## Conclusion

This study has tried to understand and figure out the differentially expressed genes involved in the amniotic fluid RNA expression profiling for Turner's syndrome. A gene-set enrichment analysis was carried out for both the upregulated and the downregulated genes of the amniotic fluid RNA expression profile for this particular genetic disorder. Hopefully, such information obtained from this study will help us to better understand the causes behind this disease, affecting so many people all over the world. And thus provide a more effective line of treatment or prevention for Turner's syndrome. The aim of this study was also to raise the awareness for this chromosomal abnormality called Turner's syndrome among the general public and hopefully they will become more aware after knowing of this clinical investigation.



## References

- Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)*. WileyRoyal Statistical Society. <https://doi.org/10.2307/2346101>.
- Chowdhury, Saikat, and Ram Rup Sarkar. 2015. "Comparison of Human Cell Signaling Pathway Databases—evolution, Drawbacks and Challenges." *Database* 2015 (January). Oxford UniversityPress. <https://doi.org/10.1093/database/bau126>.
- Davis, S., and P. S. Meltzer. 2007. "GEOquery: A Bridge between the Gene Expression Omnibus (GEO) and BioConductor." *Bioinformatics* 23 (14): 1846–47. <https://doi.org/10.1093/bioinformatics/btm254>.
- "Fetology: Diagnosis and Management of the Fetal Patient : Mary E. D' Alton : 9780071442015." n.d. Accessed October 9, 2018. <https://www.bookdepository.com/Fetology-Diagnosis-Management-Fetal-Patient-Mary-E-D-Alton/9780071442015>.
- Gravholt, Claus Højbjerg. 2004. "Epidemiological, Endocrine and Metabolic Features in Turner Syndrome." *Undefined*. <https://www.semanticscholar.org/paper/Epidemiological%2C-endocrine-and-metabolic-features-Gravholt/8c490d9863f5a005f4adf6a5f9ec4ffabea3830e>.
- Hui, Lisa, Donna K Slonim, Heather C Wick, Kirby L Johnson, and Diana W Bianchi. 2012. "The Amniotic Fluid Transcriptome: A Source of Novel Information about Human Fetal Development." *Obstetrics and Gynecology* 119 (1). NIH Public Access: 111–18. <https://doi.org/10.1097/AOG.0b013e31823d4150>.
- Khan, Aziz, Oriol Fornes, Arnaud Stigliani, Marius Gheorghe, Jaime A Castro-Mondragon, Robin van der Lee, Adrien Bessy, et al. 2018. "JASPAR 2018: Update of the Open-Access Database of Transcription Factor Binding Profiles and Its Web Framework." *Nucleic Acids Research* 46 (D1). Oxford University Press: D260–66. <https://doi.org/10.1093/nar/gkx1126>.
- Koide, Keiko, Donna K Slonim, Kirby L Johnson, Umadevi Tantravahi, Janet M Cowan, and Diana W Bianchi. 2011. "Transcriptomic Analysis of Cell-Free Fetal RNA Suggests a Specific Molecular Phenotype in Trisomy 18." *Human Genetics* 129 (3). NIH Public Access: 295–305. <https://doi.org/10.1007/s00439-010-0923-3>.
- Lachmann, Alexander, Denis Torre, Alexandra B. Keenan, Kathleen M. Jagodnik, Hyojin J. Lee, Moshe C. Silverstein, Lily Wang, and Avi Ma'ayan. 2017. "Massive Mining of Publicly Available RNA-Seq Data from Human and Mouse." *BioRxiv*, September. Cold Spring Harbor Laboratory, 189092. <https://doi.org/10.1101/189092>.
- Larrabee, Paige B., Kirby L. Johnson, Chaoqiang Lai, Jose Ordoas, Janet M. Cowan, Umadevi Tantravahi, and Diana W. Bianchi. 2005. "Global Gene Expression Analysis of the Living Human Fetus Using Cell-Free Messenger RNA in Amniotic Fluid." *JAMA* 293 (7). American Medical Association: 836. <https://doi.org/10.1001/jama.293.7.836>.
- Papp, Csaba, Artur Beke, Gabor Mezei, Zsanett Szigeti, Zoltán Bán, and Zoltan Papp. 2006. "Prenatal Diagnosis of Turner Syndrome." *Journal of Ultrasound in Medicine* 25 (6): 711–17. <https://doi.org/10.7863/jum.2006.25.6.711>.
- Slonim, Donna K, Keiko Koide, Kirby L Johnson, Umadevi Tantravahi, Janet M Cowan, Zina Jarrah, and Diana W Bianchi. 2009. "Functional Genomic Analysis of Amniotic Fluid Cell-Free MRNA Suggests

That Oxidative Stress Is Significant in Down Syndrome Fetuses.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (23). National Academy of Sciences: 9425–29. <https://doi.org/10.1073/pnas.0903909106>.

Smyth, Gordon K. n.d. “Limma: Linear Models for Microarray Data.” Accessed April 17, 2018. <https://pdfs.semanticscholar.org/b400/022ca3df41788bd020da732746eb7515444e.pdf>.

———. 2004. “Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments.” *Statistical Applications in Genetics and Molecular Biology* 3 (1): 1–25. <https://doi.org/10.2202/1544-6115.1027>.

Tyler, Carl, and Jennifer C Edman. 2004. “Down Syndrome, Turner Syndrome, and Klinefelter Syndrome: Primary Care throughout the Life Span.” *Primary Care* 31 (3). Elsevier: 627–48, x–xi. <https://doi.org/10.1016/j.pop.2004.04.006>.

Yu, Guangchuang, and Qing-Yu He. 2016. “ReactomePA: An R/Bioconductor Package for Reactome Pathway Analysis and Visualization.” *Molecular BioSystems* 12 (2): 477–79. <https://doi.org/10.1039/C5MB00663E>.