



UNIVERSITY OF LIFE SCIENCES "KING MIHAI I" FROM
TIMISOARA
"YOUNG PEOPLE AND MULTIDISCIPLINARY RESEARCH
IN APPLIED LIFE SCIENCES"

16-17 November 2023



"Young people and multidisciplinary
research in applied life sciences"

An identification of metastasis regulators in chicken (Gallus gallus) sarcoma cell lines using transcriptomics data

Phuong Y Nhu Doan¹, Jiale Wang², Phoymany Somany³, Aisha Adan Jaldesa⁴, Adrienn Szarvas (mentor)

Abstract: Rous sarcoma virus (RSV), which is an oncovirus, can cause sarcoma and consequently induce malignant tumors in chicken (Gallus gallus). Research into molecular factors that regulate the tumor-inducing ability are essential to develop prevention and curation methods against RSV. In this study, we aimed to determine candidate genes contributing to the formation of tumors through a transcriptomic analysis in R programming with GSE42516 and GSE15141, which are microarray expression dataset in GEO-NCBI database. We conducted differential expression analysis among a total of 8 metastatic samples and 5 non-metastatic samples, starting from data normalization, then creating model matrixes for pairwise comparisons and using eBayes function to calculate the log fold change values and significance level of all genes (p-value).

• Introduction

• High throughput next-generation sequencing (NGS) have been an effective tool in discovering genetic factors of metastasis and developing therapeutic strategies against tumors. Recent studies have successfully identified several regulators associated with metastasis progress in chicken sarcoma cell lines. A set of 176 genes were commonly expressed in embryo fibroblasts and neurotinal cells in chickens that were introduced with *v-src* Kovárová D. et al (2013) found that HOPX (homeodomain only protein X) were particularly down-regulated in *v-src* transformed cell lines compared to non-transformed clones Investigation into EGR1, a transcription factor, was shown to be specifically expressed in PR9692 metastasizing cell line

• In this study, we take advantage of transcriptomic datas from previous two studies to predict a list of candidate metastasis regulators that play vital roles in tumor formation of chicken. We conducted an in-depth bioinformatics analysis to identify a set of differentially expressed genes (DEGs) by using R Bioconductor packages across all experimental samples, starting from data normalization, then creating model matrixes for pairwise comparisons and using eBayes function to calculate significance level (p-value) and log fold change values (logFC) of all genes. The identified genes may have a pivotal role in elucidating molecular mechanisms and developing effective strategies against sarcoma virus.

• Material and method

1. Data collection:



GSE42516 (2 metastasis vs 2 non-metastasis)

GSE15141 (6 metastasis vs 3 non-metastasis)

2. Quantile normalization

The expression level of retrieved samples were subjected to $\log_2$ transformation

```
##normalize gene expression level  
summary(exprs(gse36272))  
exprs(gse36272) <- log2(exprs(gse36272))
```

3. Differential Expression Analysis

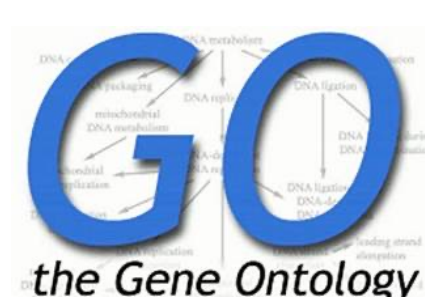
Limma (Linear Model for MicroArray Data) package were implemented in Rstudio

```
##create model matrix  
library(limma)  
design <- model.matrix(~0+sampleinfo$sel)  
fit <- lmFit(exprs(gse36272), levels=design)  
head(fit$coefficients)
```

```
##pairwise comparisons  
contrasts <- makeContrasts(Xoo - control, levels = design)  
fit2 <- contrasts.fit(contrasts, fit)  
fit2 <- eBayes(fit2)  
decideTests(fit2)  
table(decideTests(fit2))
```

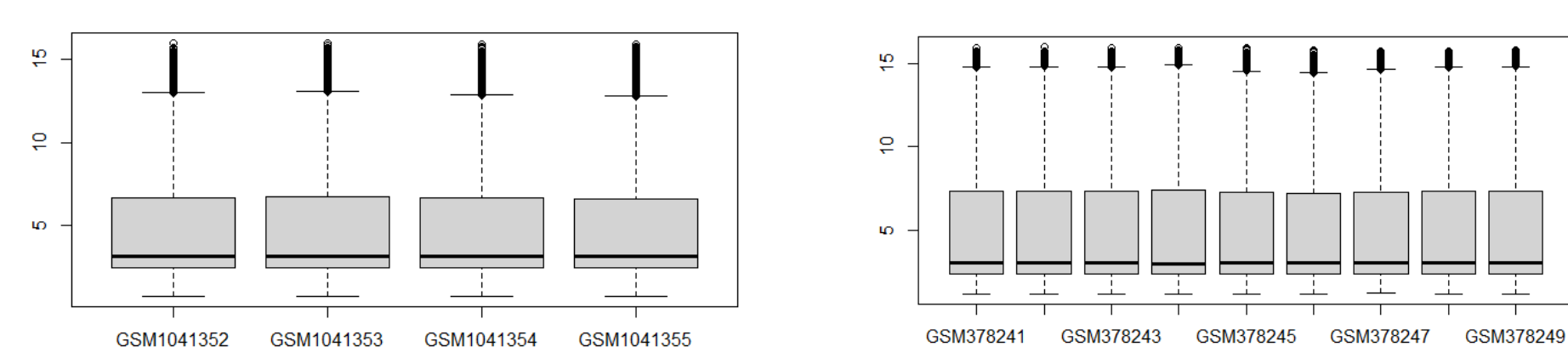
4. Enrichment analysis

Significant DEGs were submitted to DAVID tool for Gene Ontology (GO) and KEGG Pathway Enrichment analysis



• Results and discussions

1. Data normalization



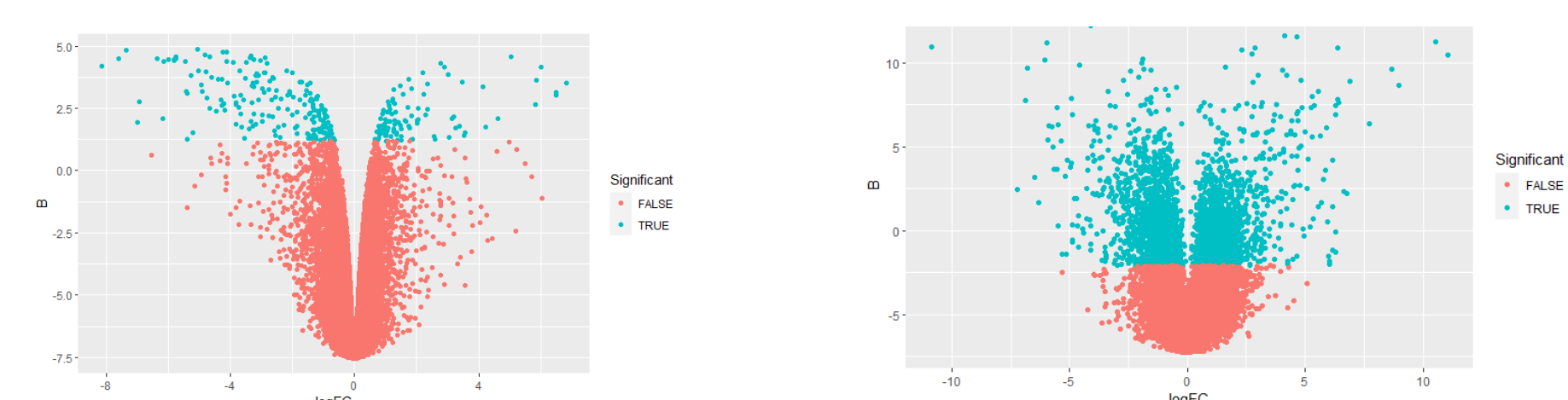
(A)

(B)

Figure 1. (A) Expression level of 4 samples (GSM1041352, GSM1041353, GSM1041354, GSM1041355) from GSE42516 (maximum expression values are not higher than 16)

(B) Gene expression level of 9 samples from GSE15141 (maximum expression values are not higher than 16)

2. Differential expression analysis



(A)

(B)

Figure 2 A. Differential expression analysis of GSE42516, including insignificant DEGs with p-value ≥ 0.05 (indicated by red dots) and significant DEGs with p-value ≤ 0.05 (indicated by blue dots) (genes that have $\log_2FC \geq 1$ are considered upregulated, genes with $\log_2FC \leq -1$ are considered downregulated). Differential expression analysis of GSE15141.

(B) A total of 1444 downregulated genes and 1314 regulated genes were identified among significant DEGs

3. Enrichment analysis

4. BP term	Gene count	Percent (%)	p-value	Benjamini
Heterochromatin assembly	10	0.9	1.2e-4	2.6e-1
Negative regulation of megakaryocyte differentiation	5	0.4	2.4e-3	1.0e
Endocytosis	14	1.3	2.8e-3	1.0e
DNA replication – independent nucleosome assembly	6	0.5	2.9e-3	1.0e
Endosomal transport	8	0.7	8.7e-3	1.0e
Positive regulation of cell differentiation	4	0.4	1.0e-2	1.0e
Positive regulation of mitochondrial fission	5	0.4	1.1e-2	1.0e
DNA replication – dependent nucleosome assembly	5	0.4	1.4e-2	1.0e
Secretory granule organization	3	0.3	1.4e-2	1.0e

Figure 3. Table of top Biological Process that are mainly enriched in significant DEGs list

• Conclusions

We demonstrated a typical workflow of in silico analysis to identify key genes across metastasis samples in published transcriptomics data. The list of potential candidate metastasis regulators were provided as top most significant differentially expressed genes between metastasis and non-metastasis samples. These molecular regulators can be utilized for early diagnosis of sarcoma virus infection, as well as effective therapeutic strategies against avian leukosis. However, further in vitro experiments are necessary to confirm the activity and expression of discovered genes.